

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
7. November 2002 (07.11.2002)

PCT

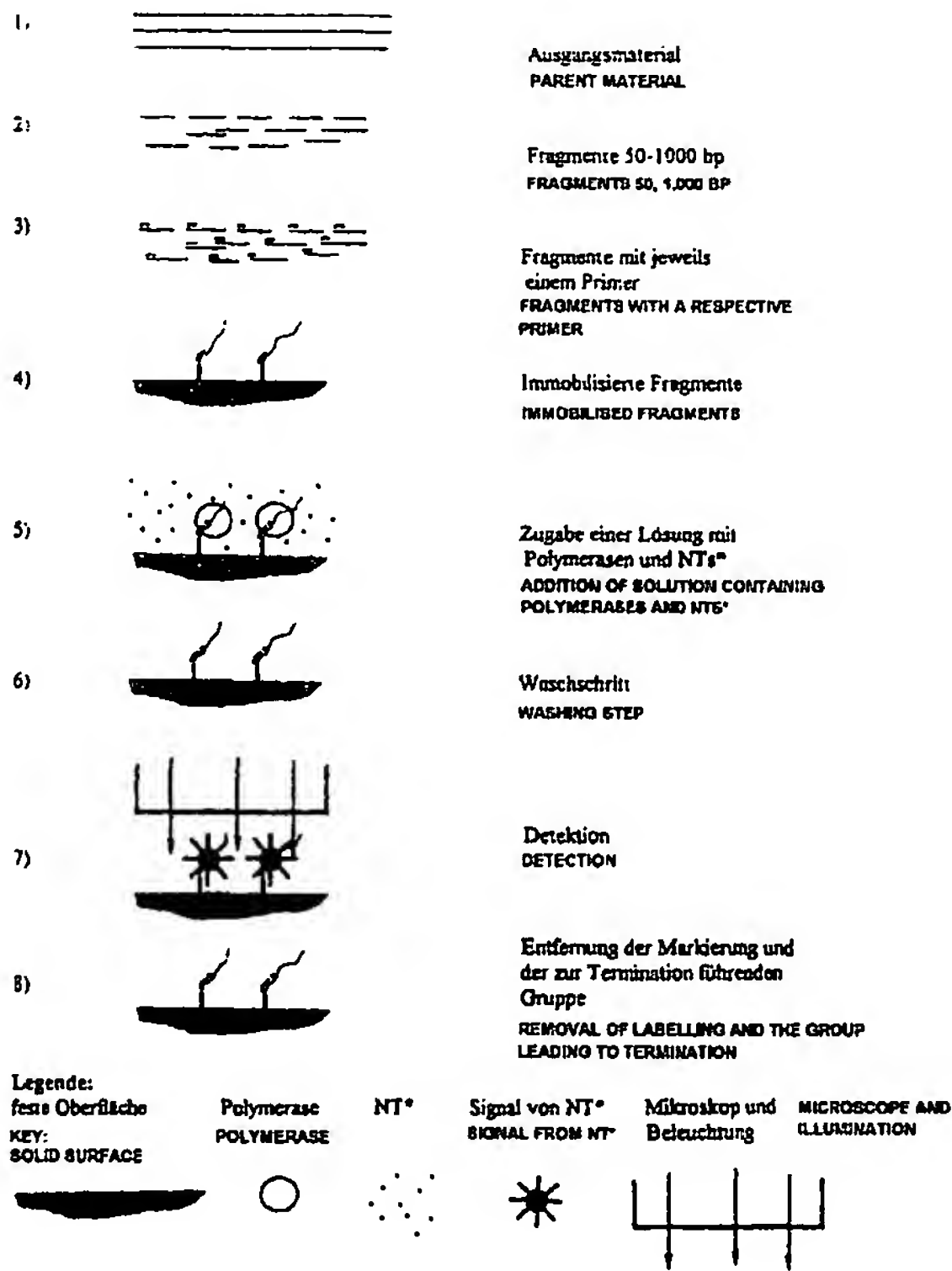
(10) Internationale Veröffentlichungsnummer
WO 02/088381 A2

- (51) Internationale Patentklassifikation⁷: C12Q 1/00 (72) Erfinder; und
(21) Internationales Aktenzeichen: PCT/EP02/04657 (75) Erfinder/Anmelder (nur für US): TCHERKASSOV,
Dmitri [RU/DE]; Kahlhorststrasse 36, 23562 Lübeck
(DE).
(22) Internationales Anmeldedatum: 26. April 2002 (26.04.2002) (74) Anwälte: WEBER-QUITZAU, Martin usw.; Uexküll &
Stolberg, Beselerstr. 4, 22607 Hamburg (DE).
(25) Einreichungssprache: Deutsch
(26) Veröffentlichungssprache: Deutsch (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VN, YU, ZA, ZM, ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR DETERMINING GENE EXPRESSION

(54) Bezeichnung: VERFAHREN ZUR BESTIMMUNG DER GENEXPRESSSION



(57) Abstract: The invention relates to a method for analyzing gene expression, i.e. for parallel analysis of the expression of a large number of genes. Said method is based on analysis of sequences on a large number of bound chains. Short sequence sections from each of these chains are determined, followed by evaluation and comparison with gene sequences in data banks, providing information about the genes thus expressed and the intensity of said expression.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Analyse der Genexpression, d.h. zur parallelen Analyse der Expression einer grossen Anzahl von Genen. Die Methode basiert auf der Analyse von Sequenzen an vielen gebundenen Nukleinsäureketten. Dazu werden kurze Sequenzabschnitte aus jeder dieser Ketten ermittelt. Die anschliessende Auswertung und der Vergleich mit Gensequenzen in Datenbanken erlaubt Aussagen über die exprimierten Gene und die Stärke ihrer Expression.

BEST AVAILABLE COPY

WO 02/088381 A2



(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

Verfahren zur Bestimmung der Genexpression

Zusammenfassung:

Die Erfindung beschreibt ein Verfahren zur parallelen Analyse der Expression einer großen Anzahl von Genen. Diese Methode basiert auf der Analyse von Sequenzen an vielen gebundenen Nukleinsäureketten. Dazu werden kurze Sequenzabschnitte aus jeder dieser Ketten ermittelt. Die anschließende Auswertung und der Vergleich mit Gensequenzen in Datenbanken erlaubt Aussagen über die exprimierten Gene und die Stärke ihrer Expression.

1. Abkürzungen und Begriffserläuterungen

DNA - Desoxyribonukleinsäure verschiedenen Ursprungs und unterschiedlicher Länge: (genomische DNA, cDNA, ssDNA, dsDNA)

RNA - Ribonukleinsäure (meist mRNA)

Genprodukte - mRNA-Transkripte oder von der mRNA abgeleitete Nukleinsäureketten (z.B. einzelsträngige cDNAs, doppelsträngige cDNAs, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA).

cDNA - komplementäre DNA

10 Polymerasen - Enzyme, die komplementäre Nukleotide in einen wachsenden DNA- oder RNA-Strang einbauen können (z.B. DNA-Polymerasen, Reverse-Transcriptasen, RNA-Polymerasen)

dNTP - 2'-deoxy-Nucleosid-Triphosphate, Substrate für DNA-
15 Polymerasen und Reverse-Transcriptasen

NTP - Nukleosid-Triphosphate, Substrate für RNA-Polymerasen

NT - natürliches Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet.

20

Abkürzung "NT" wird auch bei der Längenangabe einer Nukleinsäuresequenz verwendet, z.B. 1.000 NT. In diesem Fall steht "NT" für Nukleosid-Monophosphate.

25 Im Text wird bei Abkürzungen die Mehrzahl durch Verwendung des Suffixes "s" gebildet, "NT" steht zum Beispiel für "Nukleotid", "NTs" steht für mehrere Nukleotide.

NT' - modifiziertes Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet. NTs' bedeutet: modifizierte Nukleotide
30

Plane Oberfläche: Oberfläche, die vorzugsweise folgende Merkmale aufweist: 1) Sie erlaubt, mehrere einzelne Moleküle, vorzugsweise mehr als 100, noch besser mehr als 1000, mit dem jeweiligen gegebenen Objektiv-Oberfläche-Abstand bei einer Objektivposition gleichzeitig zu detektieren. 2) Die
35

immobilisierten bzw. gebundenen einzelnen Moleküle befinden sich in derselben Fokusebene, die reproduzierbar eingestellt werden kann.

- 5 Sterisches Hindernis: Sterisch anspruchsvolle Gruppe, die durch ihre chemische Struktur die Eigenschaften der mit dieser Gruppe gekoppelten NTs' so verändert, dass diese durch eine Polymerase in einer Extensionsreaktion nicht nacheinander eingebaut werden können.

10

Definition der Termination: Als Termination wird in dieser Anmeldung der reversible Stop des Einbaus der modifizierten ungespalteten NTs' bezeichnet.

Dieser Begriff ist von dem üblichen Gebrauch des Wortes "Termination" durch Dideoxy-NTP bei einer konventionellen Sequenzierung zu trennen.

Die Termination kommt nach dem Einbau eines modifizierten NT' zustande. Das modifizierte eingebaute NT' trägt eine an die Base reversibel gekoppelte sterische Gruppe, die zur Behinderung des
20 Einbaus eines nächsten komplementären NT' in den wachsenden Strang durch eine Polymerase führt.

SNP - single nucleotide polymorphism

25 PBS - Primerbindungsstelle

Weitfeld-Optik-Detektionssystem - Detektionssystem, das gleichzeitig Fluoreszenzsignale von einzelnen, auf einer Fläche verteilten Molekülen detektieren kann, wobei die Fläche
30 ca. 100 μm^2 und größer ist. Ein Beispiel für Weitfeld-Detektionsoptik stellt Fluoreszenzmikroskop Axiovert 200 oder Axioplan 2e (Zeiss) mit einem Planneofluar-Objektiv 100x NA 1.4 Ölimmersion (Zeiss), oder einem Planapochromat-Objektiv 100x NA 1.4 Ölimmersion (Zeiss) dar; die Anregung der
35 Fluoreszenz kann dabei mit einer Lampe, z.B. Quecksilberdampf- oder Laser oder Dioden erfolgen. Sowohl Epifluoreszenzmodus als auch Totalreflexions-

Fluoreszenzmikroskopie-Modus (total internal reflection fluorescence microscopy, TIRF-Microscopy) oder Laser-Scanning-Mikroskopie-Modus können verwendet werden. In dieser Anmeldung wird Gebrauch von dieser Weitfeld-Detektionsoptik gemacht.

5

2. Stand der Technik:

Die Analyse des Genexpressions-Spektrums ist zu einem wichtigen
10 Werkzeug in der Wissenschaft geworden. Der Vergleich der Genexpressions-Spektren zwischen verschiedenen Zelllinien, Geweben oder Entwicklungsstadien erlaubt Rückschlüsse auf die darin ablaufenden spezifischen biologischen Prozesse. So kann man z.B. erwarten, dass der Vergleich zwischen Tumorzellen und
15 gesunden Zellen gleicher Herkunft Auskunft über die am Tumorgeschehen beteiligten Gene gibt. Dabei ist wichtig, dass die Aktivität möglichst vieler oder aller Gene gleichzeitig analysiert wird.

20 Die Analyse der Genexpression ist eine komplexe Aufgabe: Die Anzahl der in einem Zelltyp aktiven Gene kann mehrere Tausend betragen. Die Analyse sollte aber möglichst alle im Genom der betreffenden Art enthaltenen Gene (etwa 32000 beim Menschen) berücksichtigen. Hinzu kommt, dass die im jeweiligen Zelltyp
25 aktiven Gene erstens meist noch nicht komplett bekannt sind und zweitens unterschiedlich stark exprimiert werden.

Es wurden bereits viele Methoden zur Genexpressionsanalyse entwickelt, so z.B. Differential Display (Nature 1984, v.308
30 S.149, Science 1992 v.257 S.967), Expressed Sequence Tags (EST) (Science 1991 v.252, S.1656, Nature Genetics 1992, v.2 S.173), Northern blotting oder RT-PCR (PNAS 1977, v.74, S.5350, Cell 1983 v.34 S.865, "The PCR Technique, RT-PCR" 1998, Ed. Paul Suebert, Eaton Publishing). Alle diese Methoden können nur eine
35 sehr begrenzte Anzahl an Genen pro Reaktion analysieren und sind zum Teil sehr arbeitsintensiv.

Die am weitesten verbreitete Methode zur parallelen Analyse der Genexpressionsmuster ist die Hybridisierung eines zu analysierenden Gemischs von cDNA-Molekülen mit an eine Oberfläche gebundenen Oligonukleotiden, die in einer bestimmten Anordnung, meist als "Microarray" fixiert sind ("Microarray Biochip Technology" 2000, Ed. M.Schena, Eaton Publishing, Zhao et al. Gene 1995, v. 156, S.207, Schena et al. Science 1995 v.270, S.467, Lockhart et al. US Patent 6.040.138, Wang US Patent 6.004.755, Arlinghaus et al. US Patent 5.821.060, 10 Southern US Patent 5.700.637, Fodor et al. US Patent 5.871.928).

Zu den großen Nachteilen der Hybridisierungsmethode zählen: Die Fertigung der an die Oberfläche gebundenen Oligonukleotide ist 15 teuer. Die Analyse beschränkt sich auf Gene, deren Sequenzen bereits bekannt sind. Mehrere Mismatch-Kontrollen vergrößern die Anzahl der Oligonukleotide, die immobilisiert werden müssen.

20 Die vorliegende Erfindung beschreibt eine Methode zur parallelen Analyse einer großen Anzahl von Nukleinsäurekettenmolekülen. Sie basiert auf der Detektion der Fluoreszenzsignale von einzelnen Molekülen. Durch eine gleichzeitige Sequenzierung einzelner Genproduktmoleküle erhält 25 die Methode mehrere Vorteile gegenüber dem Stand der Technik:

1) Die Genprodukte können in einer beliebigen Anordnung auf der Oberfläche binden. Eine vorherige aufwendige Synthese von verschiedenen Oligonukleotiden an bestimmten Positionen (wie 30 beispielsweise bei der Hybridisierungsmethode) ist somit nicht notwendig.

2) Das Material kann auf einer standardisierten Oberfläche analysiert werden.

35

3) Auch die Expression noch unbekannter Gene kann ermittelt werden, weil alle im Ansatz enthaltenen Genprodukte analysiert

werden. Dies ist mit konventionellen DNA-Arrays nicht möglich.

4) Die große Anzahl der analysierten Moleküle erlaubt auch die Detektion schwach exprimierter Gene.

5

5) Kleinste Mengen an Ausgangsmaterial können eingesetzt werden: mRNA aus einer einzelnen Zelle kann für die Analyse ausreichend sein.

10 6) Sämtliche Schritte des Verfahrens können weitgehend automatisiert werden.

3. Kurze Beschreibung:

15

Die Erfindung beschreibt ein Verfahren zur parallelen Analyse der Genexpression, bei dem man

einzelsträngige Genprodukte bereitstellt, man

20

die Genprodukte unter Verwendung eines einheitlichen oder mehrerer unterschiedlicher Primer in Form von Genprodukt-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

25

eine zyklische Aufbaureaktion des komplementären Stranges der Genprodukte unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

30 a) zu den auf der Oberfläche gebundenen Genprodukt-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens
35 zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale vonein-

- ander unterscheiden lassen, wobei die NTs* strukturell an der Base so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT* in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist, man
- 5
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT* verlängert werden, man
- 10
- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs* geeignet sind, man
- 15
- d) die einzelnen, in komplementäre Stränge eingebauten NTs* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- 20
- e) zur Erzeugung unmarkierter, (NTs oder) Genprodukte die Fluoreszenzfarbstoffe und die sterisch anspruchsvollen Liganden von den am komplementären Strang angefügten NTs* abspaltet, man
- 25
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man
- 30

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

35 wobei man die relative Position einzelner Genprodukt-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser Genprodukte durch spezifische Zuordnung der in Stufe d) in auf-

einanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt und man aus den ermittelten Teilsequenzen die Identität der Genprodukte bestimmt.

5

Bei den Genprodukten handelt es sich um die primären Genprodukte der Gene, deren Expression analysiert werden soll. Im wesentlichen handelt es sich dabei um RNA-Transkripte der genannten Gene, welche auch als Target-Sequenzen (oder Target-
10 Nukleinsäuresequenzen) bezeichnet werden. Diese Target-Sequenzen schließen neben mRNA auch davon abgeleitete einzelsträngige und doppelsträngige cDNA, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA ein.

15 Die Genprodukte oder Target-Sequenzen können entweder als mRNAs direkt aus einer biologischen Probe (z.B. Zellextrakt, Gewebeextrakt oder Extrakt von ganzen Organismen) isoliert oder als cDNAs durch reverse Transkription der mRNAs erhalten werden.

20

Unter einer hoch parallelen Analyse wird in diesem Zusammenhang die gleichzeitige Sequenzanalyse vieler Genprodukt-Moleküle verstanden (beispielsweise mindestens etwa 100.000, vorzugsweise mindestens etwa 500.000, insbesondere aber
25 mindestens etwa 1.000.000 bis etwa 10.000.000), wobei diese Genprodukt-Moleküle eine komplexe heterogene Population darstellen, die z.B. einem kompletten Expressionsprofil bzw. einem Expressionsspektrum eines Gewebes entspricht.

30 Gemäß einer besonderen Ausführungsform der Erfindung kann das Verfahren durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man

a) in jedem Zyklus nur jeweils ein markiertes NT',

35 b) in jedem Zyklus jeweils zwei unterschiedlich markierte NTs' oder

c) in jedem Zyklus jeweils vier unterschiedlich markierte NTs'

einsetzt.

Das Verfahren kann auch durchgeführt werden, indem man die
5 Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach
wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei
unterschiedlich markierte NTs' und zwei unmarkierte NTs einsetzt
und man die Identität der Genprodukte durch Vergleich mit den
Referenzsequenzen ermittelt.

10

Gegenstand der vorliegenden Erfindung sind ferner die in den
Figuren 6e, 6f und 6g dargestellten Nukleotide und die ent-
sprechenden markierten Nukleotide, die beispielsweise an die
terminale Aminofunktion angeheftete Fluoreszenzfarbstoffe
15 aufweisen, oder die in den Figuren 6h, 6i oder 6j dargestellten
markierten Nukleotide.

Gegenstand der vorliegenden Erfindung ist ferner die Verwendung
der in Figuren 6e, 6f und 6g dargestellten Nukleotide und der
20 entsprechenden mit einem Fluoreszenzfarbstoff markierten
Nukleotide für das erfindungsgemäßen Verfahren.

Gegenstand der vorliegenden Erfindung ist ferner die Verwendung
der an der Base modifizierten NT*s (Beispiele siehe Figuren
25 6k, 6L und 6m) und der entsprechenden mit einem
Fluoreszenzfarbstoff markierten Nukleotide für das
erfindungsgemäße Verfahren.

Die Methode basiert auf mehreren Prinzipien:

30

1. Kurze Nukleotidsequenzen (10-50 NTs) enthalten genügend
Informationen zur Identifizierung des korrespondierenden Gens,
wenn die Gensequenz selbst bereits in einer Datenbank enthalten
ist.

35 Eine Sequenz aus beispielsweise 10 NTs kann mehr als 10^6
verschiedene Kombinationen bilden. Das ist z.B. für die meisten
Gene im menschlichen Genom, das nach heutiger Schätzung 32000

Gene enthält, ausreichend. Für Organismen mit weniger Genen kann die Sequenz noch kürzer sein.

2. Der Methode liegt ein neues Verfahren zur Sequenzierung einzelner Nukleinsäurekettenmoleküle zugrunde.

3. Es können Nukleinsäureketten-Gemische untersucht werden.

4. Die Sequenzierungsreaktion läuft an vielen Molekülen gleichzeitig ab, wobei die Sequenz jeder einzelnen gebundenen Nukleinsäurekette analysiert wird.

Es ist bekannt, dass zur Untersuchung der Genexpression mRNAs oder von der mRNA abgeleitete Nukleinsäureketten (z.B. einzelsträngige cDNAs, doppelsträngige cDNAs, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA) eingesetzt werden kann. Unabhängig von der genauen Zusammensetzung werden sie im folgenden als Genprodukte bezeichnet. Auch Teilsequenzen dieser Genprodukte werden im folgenden als Genprodukte bezeichnet.

Diese Genprodukte stellen ein Gemisch aus verschiedenen Nukleinsäureketten dar.

Als Grundlage der Analyse dient die Synthese eines zum Genprodukt komplementären Stranges.

Das Ziel der Vorbereitung ist, auf einer planen Oberfläche in zufälliger Weise gebundene Genprodukt-Primer-Komplexe bereitzustellen, an denen der Einbau von NT's durch die Polymerase stattfinden kann (extensionsfähige Genprodukt-Primer-Komplexe).

Mit diesen gebundenen Genprodukt-Primer-Komplexen wird die Sequenzierungsreaktion durchgeführt.

Sie verläuft in mehreren Zyklen. Pro Zyklus wird jeweils nur ein einziges markiertes NT' in den wachsenden Strang eingebaut.

Ein Zyklus umfasst folgende Schritte:

a) Zugabe einer Lösung mit markierten Nukleotiden (NTs') und Polymerase zu gebundenen Genprodukt-Primer-Komplexen,

- b) Inkubation der gebundenen Genprodukt-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
 - c) Waschen,
 - 5 d) Detektion der Signale von einzelnen modifizierten, in die neu synthetisierten Stränge eingebauten NTs'-Molekülen,
 - e) Entfernung der Markierung von den eingebauten Nukleotiden,
 - f) Waschen.
- 10 Dieser Zyklus kann mehrmals wiederholt werden, so dass von jedem an der Sequenzierungsreaktion teilnehmenden Genprodukt-Primer-Komplex vorzugsweise 10 bis 50 NTs ermittelt werden. Danach erfolgt die Rekonstruktion der Nukleinsäuresequenzen aus den detektierten Signalen. Die ermittelten Sequenzen der
- 15 gebundenen Genprodukte werden zur Bestimmung der Abundanzen untereinander verglichen und durch Vergleich mit Gensequenzen in Datenbanken bestimmten Genen zugeordnet.

Diese allgemeinen Prinzipien der Reaktion werden anhand eines

20 Beispiels erklärt (Fig.1):

Als Ausgangsmaterial dient ein Gemisch aus mRNA-Molekülen (Genprodukte). Das zu analysierende Material (1) wird auf einer geeigneten planen Oberfläche gebunden (2). Dabei vermitteln die

25 zuvor auf der Oberfläche gebundenen Oligo-dT-Primer die Bindung von Genprodukten an die Oberfläche. Die Genprodukte werden durch Hybridisierung (Annealing) an die Primer gebunden. Die nicht gebundenen mRNA-Moleküle werden durch einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der

30 gesamten Reaktionsoberfläche durchgeführt, wobei in die extensionsfähigen Genprodukt-Primer-Komplexe markierte NT's eingebaut werden. Diese Reaktion verläuft zyklisch. Im 1. Schritt des Zyklus wird ein markiertes NT' in den wachsenden Strang eingebaut: Dabei wird die Reaktion so gesteuert, dass in

35 jedem Zyklus jeweils nur ein markiertes NT' von einer Polymerase (Reverse Transcriptase) in den wachsenden Strang eingebaut werden kann. Das wird durch die Verwendung von NTs' erreicht,

die eine reversibel gekoppelte, zur Termination führende Gruppe tragen. Der Einbau eines weiteren markierten NT' wird dadurch verhindert. Die Polymerase und die markierten NTs' werden gleichzeitig in die Reaktion eingesetzt (3). Danach wird das
5 Reaktionsgemisch entfernt und die Oberfläche in geeigneter Art und Weise gewaschen (4). Nun folgt ein Detektionsschritt (5). Die Oberfläche wird mit einer für die Einzelmoleküldetektion geeigneten Vorrichtung (bestehend z.B. aus Lichtquelle, Mikroskop, Kamera, Scantisch, Computer mit Steuerungs- und Bild-
10 erkennungs- bzw. Bildverarbeitungssoftware) abgescannt und die Signale der einzelnen, eingebauten markierten NTs' identifiziert. Nach dem Detektionsschritt wird die Markierung und die zur Termination führende Gruppe von allen eingebauten NTs' entfernt (6). Die Entfernung kann chemisch, physikalisch
15 oder enzymatisch erfolgen. Nach einem sich anschließenden Waschschrift kann ein neuer Zyklus beginnen.

4. Detaillierte Beschreibung

20

4.1 Auswahl des Materials:

Genprodukte können von verschiedenen biologischen Objekten stammen, so z.B. von einzelnen Zellen, Zellpopulationen, einem
25 Gewebe oder von kompletten Organismen. Auch biologische Flüssigkeiten wie Blut, Sputum oder Liquor können als Quelle der Genprodukte dienen. Die Methoden zur Gewinnung der Genprodukte aus den verschiedenen biologischen Objekten sind beispielsweise folgenden Literaturquellen zu entnehmen:
30 "Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory , "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc..

Es kann sowohl die Gesamtheit der isolierten Genprodukte als
35 auch ein durch eine Vorselektion ausgewählter Teil davon in die Sequenzierungsreaktion eingesetzt werden. Durch Vorselektion kann man die Menge der zu analysierenden Genprodukte

reduzieren. Die Vorselektion kann beispielsweise durch molekularbiologische Verfahren wie z.B. PCR-Amplifikation, Gel-Auftrennung oder Hybridisierung mit anderen Nukleinsäureketten erfolgen ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.) Vorzugsweise wird die Gesamtheit der Genprodukte als Ausgangsmaterial gewählt.

10 4.2 Vorbereitung des Materials:

Ziel der Vorbereitung des Materials ist es, aus dem Ausgangsmaterial an die Oberfläche gebundene, extensionsfähige Genprodukt-Primer-Komplexe zu bilden. Diese haben dann vorzugsweise die in Fig. 2 dargestellte Struktur. Wobei pro Genprodukt 15 maximal nur ein Primer binden sollte.

4.2.1 Primerbindungsstelle (PBS):

Jedes Genprodukt hat vorzugsweise nur eine Primerbindungsstelle.

20 Eine Primerbindungsstelle ist ein Sequenzabschnitt, der eine selektive Bindung des Primers an das Genprodukt ermöglichen soll.

Als Primerbindungsstellen können Abschnitte in der 25 Nukleinsäuresequenz dienen, die in den zu analysierenden Sequenzen natürlicherweise vorkommen (z.B. polyA-Strecken in mRNA). Eine Primerbindungsstelle kann auch zusätzlich in das Genprodukt eingeführt werden ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in 30 Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.).

Aus Gründen der Vereinfachung der Analyse kann es wichtig sein, dass eine möglichst einheitliche Primerbindungsstelle in allen 35 Genprodukten vorhanden ist. Dann können Primer mit einheitlicher Struktur in die Reaktion eingesetzt werden. Die Zusammensetzung der Primerbindungsstelle ist nicht einge-

schränkt. Ihre Länge beträgt vorzugsweise zwischen 10 und 100 NTs. Die Primerbindungsstelle kann eine funktionelle Gruppe tragen, beispielsweise zur Bindung des Genprodukts an die Oberfläche. Diese funktionelle Gruppe kann z.B. eine Biotin-5 oder Digoxigenin-Gruppe sein.

Als Beispiel für die Einführung einer Primerbindungsstelle in die Genprodukte wird das Nukleotid-Tailing von antisense cDNA-Fragmenten beschrieben.

10

Als erstes werden einzelsträngige cDNAs von mRNAs synthetisiert. Es resultiert eine Population an cDNA-Molekülen, die eine Kopie der mRNA-Population darstellen, sogenannte antisense-cDNA. (Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.). Mit einer terminalen Deoxynucleotidyltransferase kann man mehrere (z.B. zwischen 10 und 20) Nukleosid-monophosphate an das 3'-Ende dieser antisense cDNA anknüpfen, z.B. mehrere 20 Adenosin-Monophosphate ((dA)n-Tail genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (dT)n-Primers, verwendet ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38) (Fig. 3).

25

4.2.2 Primer für die Sequenzierungsreaktion

Dieser hat die Funktion, den Start an einer einzigen Stelle des Genprodukts zu ermöglichen. Vorzugsweise bindet er an die Primerbindungsstelle im Genprodukt. Die Zusammensetzung und die 30 Länge des Primers sind nicht eingeschränkt. Außer der Startfunktion kann der Primer auch andere Funktionen übernehmen, wie z.B. eine Verbindung der Genprodukt-Primer-Komplexe zur Reaktionsoberfläche zu schaffen. Primer sollten so an die Länge und Zusammensetzung der Primerbindungsstelle 35 angepaßt werden, dass der Primer den Start der Sequenzierungsreaktion mit der jeweiligen Polymerase ermöglicht.

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 15 und 30 NTs. Der Primer kann eine funktionelle Gruppe tragen, die beispielsweise zur Bindung des Primers an die Oberfläche dient, beispielsweise ist eine
5 solche funktionelle Gruppe eine Biotingruppe (s. Abschnitt Immobilisierung). Sie soll die Sequenzierung nicht stören. Die Synthese eines solchen Primers kann z.B. mit dem DNA-Synthesizer 380 A Applied Biosystems ausgeführt werden oder aber als Auftragssynthese bei einem kommerziellen Anbieter,
10 z.B. MWG-Biotech GmbH, Deutschland, erstellt werden.

Es können auch unterschiedliche Primer verwendet werden, ein definierter Primersatz, oder ein Primergemisch.

15 Der Primer kann vor der Hybridisierung an die zu analysierenden Fragmente auf der Oberfläche mit verschiedenen Techniken fixiert oder direkt auf der Oberfläche synthetisiert werden beispielsweise nach (McGall et al. US Patent 5412087, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734,
20 "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Fodor et al. Science 1991 v.285 S.767, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR
25 1992 v.20 S.1679).

Die Primer werden auf der Oberfläche in einer Dichte zwischen 10 bis 100 pro 100 μm^2 , 100 bis 10.000 pro 100 μm^2 , 10.000 bis 1.000.000 pro 100 μm^2 oder größer als 1.000.000 pro 100 μm^2
30 gebunden.

Der Primer oder das Primergemisch wird mit Genprodukten unter Hybridisierungsbedingungen inkubiert, die ihn selektiv an die Primerbindungsstelle jedes Genprodukts binden lassen. Diese
35 Primer-Hybridisierung (Annealing) kann vor (1), während (2) oder nach (3) der Bindung der Genprodukte an die Oberfläche erfolgen. Falls Genprodukte als doppelsträngige Nukleinsäuren

vorliegen, werden sie vor der Hybridisierung durch Hitze denaturiert ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press). Die Optimierung der Hybridisierungsbedingungen hängt von der genauen Struktur der Primerbindungsstelle und des Primers ab und läßt sich nach Rychlik et al. (NAR 1990 v.18 S.6409) berechnen. Im folgenden werden diese Hybridisierungsbedingungen als standardisierte Hybridisierungsbedingungen bezeichnet.

- 10 Falls das Ausgangsmaterial eine poly-A-Strecke oder eine poly-dA-Strecke aufweist (z.B. mRNA, sense cDNA oder antisense-cDNA mit $(dA)_n$ -Tail) kann man einen oligo-dT-Primer verwenden. Es kann allerdings auch ein Primergemisch bestehend aus 12 verschiedenen Primern mit folgender allgemeiner Struktur
15 5' $(K)_n MN3'$ verwendet werden. Wobei (n) zwischen 10 und 50 liegt, vorzugsweise zwischen 20 und 30. "K" steht für dT oder dU, "M" und "N" stehen jeweils für dA, dT oder dU, dC, dG (z.B. 5'-dTdTdTdTdTdTdTdTdT₁₀dTdTdTdTdTdTdTdTdT₂₀dAdG-3'). Ein solches Primergemisch ermöglicht einen exakten Start der
20 Sequenzierungsreaktion am Ende der polyA-Strecke oder der poly-dA-Strecke (geankerter Primer).

4.2.3 Fixierung von Genprodukt-Primer-Komplexen an die Oberfläche (Bindung bzw. Immobilisierung von Genprodukten).

25

- Ziel der Fixierung (Bindung, Immobilisierung) ist es, Genprodukt-Primer-Komplexe auf einer geeigneten planen Oberfläche in einer Art und Weise zu fixieren, dass eine zyklische enzymatische Sequenzierungsreaktion ablaufen kann.
30 Dies kann beispielsweise durch Bindung des Primers (s.o.) oder des Genprodukts an die Oberfläche erfolgen.

Die Reihenfolge der Schritte bei der Bindung von Genprodukt-Primer-Komplexen kann variabel sein:

- 35 1) Die Genprodukt-Primer-Komplexe können zunächst in einer Lösung durch Hybridisierung (Annealing) gebildet und anschließend an die Oberfläche gebunden werden.

- 2) Primer können zunächst auf einer Oberfläche gebunden werden und Genprodukte anschließend an die gebundenen Primer hybridisiert werden, wobei Genprodukt-Primer-Komplexe entstehen (Genprodukte indirekt an die Oberfläche gebunden)
- 5 3) Die Genprodukte können zunächst an die Oberfläche gebunden werden (Genprodukte direkt an die Oberfläche gebunden) und im anschließenden Schritt die Primer an die gebundenen Genprodukte hybridisiert werden, wobei Genprodukt-Primer-Komplexe entstehen.
- 10 Die Immobilisierung der Genprodukte an die Oberfläche kann daher durch direkte oder indirekte Bindung erfolgen.

Oberfläche und Reaktionsoberfläche sind in dieser Anmeldung als gleichwertige Begriffe aufzufassen, außer wenn explizit auf
15 eine andere Bedeutung hingewiesen wird. Als Reaktionsoberfläche dient die Oberfläche einer festen Phase eines beliebigen Materials. Dieses Material ist vorzugsweise enzymatischen Reaktionen gegenüber inert und verursacht keine Störungen der Detektion. Silicon, Glas, Keramik, Kunststoff (z.B.
20 Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material, das diesen funktionalen Anforderungen genügt, kann verwendet werden. Vorzugsweise ist die Oberfläche nicht verformbar, denn sonst ist mit einer Verzerrung der Signale bei der wiederholten
25 Detektion zu rechnen.

Falls eine gelartige feste Phase (Oberfläche eines Gels) verwendet wird, so kann dieses Gel z.B. ein Agarose- oder Polyacrylamidgel sein. Das Gel ist vorzugsweise für Moleküle
30 mit einer Molekularmasse unter 5000 Da frei passierbar (beispielsweise kann ein 1 bis 2% Agarose-Gel oder 5 bis 15% Polyacrylamid Gel verwendet werden). Eine solche Geloberfläche hat anderen festen Oberflächen gegenüber den Vorteil, dass es zu einer wesentlich geringeren unspezifischen Bindung von NT's
35 an die Oberfläche kommt. Durch die Bindung der Genprodukt-Primer-Komplexe auf der Oberfläche ist die Detektion der Fluoreszenzsignale von eingebauten NTs' möglich. Die Signale von

freien NTs' werden nicht detektiert, weil sie nicht an das Material des Gels binden und somit nicht immobilisiert werden. Das Gel ist vorzugsweise auf einer festen Unterlage befestigt (Fig. 4). Diese feste Unterlage kann Silicon, Glas, Keramik, Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material sein.

Die Dicke des Gels beträgt vorzugsweise nicht mehr als 0,1 mm. Die Geldicke ist jedoch vorzugsweise größer als die einfache Tiefenschärfe des Objektivs, damit unspezifisch an die feste Unterlage gebundene NTs' nicht in die Fokusebene gelangen und damit detektiert werden. Wenn die Tiefenschärfe z.B. 0,3 μm beträgt, so liegt die Geldicke vorzugsweise zwischen 1 μm und 100 μm . Die Oberfläche kann als eine kontinuierliche Oberfläche oder als diskontinuierliche, aus einzelnen kleinen Bestandteilen (z.B. Agarose-Kügelchen) zusammengesetzte Oberfläche hergestellt werden (Fig. 4 a,b). Die Reaktionsoberfläche muß groß genug sein, um die notwendige Anzahl der Genprodukte bei entsprechender Dichte binden zu können. Die Reaktionsoberfläche sollte vorzugsweise nicht größer als 20 cm^2 sein.

20

Die verschiedenen Zyklusschritte erfordern einen Austausch der unterschiedlichen Reaktionslösungen über der Oberfläche. Die Reaktionsoberfläche ist vorzugsweise Bestandteil eines Reaktionsgefäßes. Das Reaktionsgefäß ist wiederum vorzugsweise Bestandteil einer Reaktionsapparatur mit Durchflußvorrichtung. Die Durchflußvorrichtung ermöglicht einen Austausch der Lösungen im Reaktionsgefäß. Der Austausch kann mit einer durch einen Computer gesteuerten Pumpvorrichtung oder manuell erfolgen. Wichtig dabei ist, dass die Oberfläche nicht austrocknet. Vorzugsweise beträgt das Volumen des Reaktionsgefäßes weniger als 50 μl . Idealerweise beträgt sein Volumen weniger als 1 μl . Ein Beispiel eines solchen Durchflußsystems ist in Fig. 5 gegeben.

35 Falls die Fixierung der Genprodukt-Primer-Komplexe auf der Oberfläche über die Genprodukte erfolgt, kann dies beispielsweise durch die Bindung der Genprodukte an einem der

beiden Ketten-Enden erfolgen. Dies kann durch entsprechende kovalente, affine oder andere Bindungen erreicht werden. Es sind viele Beispiele der Immobilisierung von Nukleinsäuren bekannt (McGall et al. US Patent 5412087, Nikiforov et al. US Patent 5610287, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Rasmussen et al. Analytical Biochemistry v.198, S.138, Allemand et al. Biophysical Journal 1997, v.73, S.2064, Trabesinger et al. Analytical Chemistry 1999, v.71, S.279, Osborne et al. Analytical Chemistry 2000, v.72, S.3678, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679). Die Fixierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der Genprodukte enthaltenden Probe auf der planen Oberfläche erreicht werden. Die Genprodukte werden auf der Oberfläche in einer Dichte zwischen 10 und 100 pro 100 μm^2 , 100 bis 10.000 pro 100 μm^2 , 10.000 bis 1000.000 pro 100 μm^2 gebunden.

Die für die Detektion notwendige Dichte von extensionsfähigen Genprodukt-Primer-Komplexen beträgt ca. 10 bis 100 pro 100 μm^2 . Sie kann vor, während oder nach der Hybridisierung der Primer an die Genprodukte erreicht werden.

Beispielhaft werden im folgenden einige Methoden zur Bindung näher dargestellt:

In einer Ausführungsform erfolgt die Bindung der Genprodukt-Primer-Komplexe über Biotin-Avidin oder Biotin-Streptavidin-Bindung. Dabei wird Avidin oder Streptavidin auf der Oberfläche kovalent gebunden, das 5'-Ende des Primers ist mit Biotin modifiziert. Nach der Hybridisierung der modifizierten Primer mit den Genprodukten (in Lösung) werden diese auf der mit Avidin/Streptavidin beschichteten Oberfläche fixiert. Die Konzentration der mit Biotin markierten Genprodukt-Primer-Komplexe sowie die Zeit der Inkubation dieser Lösung mit der Oberfläche

wird so gewählt, dass eine für die Sequenzierung geeignete Dichte erreicht wird.

In einer anderen Ausführungsform werden die für die
5 Sequenzierungsreaktion geeigneten Primer vor der
Sequenzierungsreaktion auf der Oberfläche mit geeigneten
Methoden fixiert (s.o.). Die einzelsträngigen Genprodukte mit
jeweils einer Primerbindungsstelle pro Genproduktmolekül werden
damit unter Hybridisierungsbedingungen inkubiert (Annealing).
10 Dabei binden sie an die fixierten Primer und werden dadurch an
die Oberfläche gebunden (indirekte Bindung). Die Konzentration
der einzelsträngigen Genprodukte und die
Hybridisierungsparameter (z.B. Temperatur, Zeit, Puffer) werden
so gewählt, dass man eine für die Sequenzierung geeignete
15 Dichte von ca. 10 bis 100 extensionsfähigen Genprodukt-Primer-
Komplexen pro 100 μm^2 erreicht. Nach der Hybridisierung werden
ungebundene Genprodukte durch einen Waschschrift entfernt. Bei
dieser Ausführungsform wird eine Oberfläche mit einer hohen
Primerdichte bevorzugt, z.B. ca. 1.000.000 Primer pro 100 μm^2
20 oder noch höher, da die gewünschte Dichte an Genprodukt-Primer-
Komplexen schneller erreicht wird, wobei die Genprodukte nur an
einen Teil der Primer binden.

In einer anderen Ausführungsform werden die Genprodukte an die
25 Oberfläche direkt gebunden (s.o.) und anschließend mit Primern
unter Hybridisierungsbedingungen inkubiert. Bei einer Dichte
von ca. 10 bis 100 Genprodukte pro 100 μm^2 wird man versuchen
alle verfügbaren Genprodukte mit einem Primer zu versehen und
für die Sequenzierungsreaktion verfügbar zu machen. Dies kann
30 durch hohe Primerkonzentration beispielsweise 1 bis 100 mmol/l
erreicht werden. Bei einer höheren Dichte der fixierten
Genprodukte auf der Oberfläche, beispielsweise 10.000 bis
1.000.000 pro 100 μm^2 , kann die für die optische Detektion
notwendige Dichte der Genprodukt-Primer-Komplexe während der
35 Primer-Hybridisierung erreicht werden. Dabei sind die
Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer,
Primerkonzentration) so zu wählen, dass die Primer nur an einen

Teil der immobilisierten Genprodukte binden, s. Beispiel 5. Falls die Oberfläche einer festen Phase (z.B. Silikon oder Glas) zur Immobilisation verwendet wird, wird vorzugsweise eine Blockierungslösung auf die Oberfläche vor dem Schritt (a) in 5 jedem Zyklus gebracht, die zur Vermeidung einer unspezifischen Adsorption von NTs' an der Oberfläche dient. Diese Bedingungen für eine Blockierungslösung erfüllt beispielsweise eine Albuminlösung (BSA) mit einem pH-Wert zwischen 8 und 10.

10 4.3 Wahl der Polymerase

Bei der Wahl der Polymerase spielt die Art der verwendeten immobilisierten Nukleinsäure (RNA oder DNA) eine entscheidende Rolle:

15

Falls RNA als Genprodukt (z.B. mRNA) in die Sequenzierungsreaktion eingesetzt wird, können handelsübliche RNA-abhängige DNA-Polymerasen eingesetzt werden, z.B. AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), 20 HIV-Reverse Transcriptase ohne RNase-Aktivität. Alle Reverse Transcriptasen müssen von RNase-Aktivität weitgehend frei sein ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory).

25 Falls DNA als Genprodukt (z.B. cDNA) verwendet wird, eignen sich als Polymerasen prinzipiell alle DNA-abhängigen DNA-Polymerasen ohne 3'-5' Exonuklease-Aktivität (DNA-Replication" 1992 Ed. A.Kornberg, Freeman and company NY), z.B. modifizierte T7-Polymerase vom Typ "Sequenase Version 2" (Amersham Pharmacia 30 Biotech), Klenow Fragment der DNA-Polymerase I ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech), Polymerase Beta verschiedenen Ursprungs (Animal Cell DNA Polymerases" 1983, Fry M., CRC Press Inc., kommerziell erhältlich bei Chimerx) thermostabile Polymerasen wie beispielsweise Taq- 35 Polymerase (GibcoBRL), proHA-DNA-Polymerase (Eurogentec).

4.4 Nukleotidstruktur

4.4.1 Allgemeines Prinzip

5 Für die hoch parallele Genexpressionsanalyse mit einzelnen Molekülen (parallele Sequenzanalyse von bis zu 10.000.000 Genproduktmolekülen) ist wichtig, dass jedes eingebaute NT' während der Sequenzierungsreaktion identifiziert wird. Eine Voraussetzung dafür ist, dass nur ein einziges NT' pro Zyklus in
10 die Nukleinsäurekette eingebaut wird. Falls eine Polymerase mehrere NTs' nacheinander im selben Zyklus einbaut, so führt dies zu einem Fehler in der Sequenzermittlung. Aus diesem Grund muß man den Einbau der NTs' steuern.

15 Beispielsweise wurden in der BASS-Methode reversible 3'-OH modifizierte NTs beschrieben (Dower US Patent 5.547.839, Canard et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021, Metzker et al. NAR 1994, v.22, S.4259, Welch et al. Nucleosides & Nucleotides 1999, v.18,
20 S.197). Die Spaltung soll dabei unter milden Bedingungen photochemisch (Dower US Patent 5.547.839, Welch et al. Nucleosides & Nucleotides 1999, v.18, S.197) oder chemisch (Canard et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021) erfolgen. Bei dieser Methode
25 wird eine große Anzahl gleicher einzelsträngiger DNA-Stücke an einem definierten Ort auf einer Oberfläche fixiert und das Signal von der Gesamtheit dieser vielen identischen DNA-Stücke analysiert. Zu dieser fixierten DNA wird eine Lösung mit Polymerase und Nukleotiden zugegeben, so dass ein
30 komplementärer Strang synthetisiert werden kann. Dabei soll die Polymerase schrittweise arbeiten: in jedem Schritt wird nur ein einziges Nukleotid eingebaut. Dieses wird detektiert, worauf die Polymerase in einem nächsten Zyklus das nächste Nukleotid einbaut. Bei dieser Methode wurden an der 3'-OH-Gruppe der
35 Deoxyribose modifizierte Nukleotide verwendet.

Trotz des Gelingens einiger einzelner Schritte dieser Methode

wurde sie nicht zu einem funktionsfähigen Verfahren entwickelt. Dies kann beispielsweise auf folgenden Tatsachen beruhen: Beim Aufbau der komplementären Stränge tritt sehr schnell eine Desynchronisation der Synthese auf, so dass bei jedem Schritt die Fehler akkumulieren. Deshalb können nur sehr kurze Fragmente sequenziert werden. Es ist zu betonen, dass alle beschriebenen BASS-Methoden nicht auf der Detektion von einzelnen Molekülen beruhen. Das Signal wird stattdessen von einer großen Anzahl identischer an einem definierten Ort immobilisierter Moleküle registriert. Die in diesen Methoden übliche Verwendung der Begriffe "einzelne Moleküle" und "Moleküle" zielt dabei nicht auf individuelle, voneinander getrennte Moleküle, sondern auf eine Population, die aus vielen identischen Molekülen besteht. Identisch heißt in diesem Fall, dass die Moleküle die gleiche Sequenz haben.

Ein weiteres Problem stellen die an 3'-Position modifizierte Nukleotide dar.

Die Synthese der 3'-OH-modifizierten photochemisch spaltbaren NTs ist sehr aufwendig. Die Polymerasen weisen eine sehr unterschiedliche Affinität zu diesen Nukleotidanalogen auf, so dass die Nukleinsäuresynthese sehr ungleichmäßig bzw. an vielen DNA-Stellen gar nicht abläuft (Metzker et al. NAR 1994 v.22 S.4259, Welch et al. Nucleosides & Nucleotides 1999, v.18 S.197). Aus diesem Grund eignen sich diese Analoga nicht oder nur sehr eingeschränkt für die Sequenzierungsreaktion. Eine spaltbare 3'-Ester-Verknüpfung (Canard et al. US Patent 5.798.210) kommt als Grundlage für eine reversible Termination der Synthese auch nicht in Betracht. Die meisten Polymerasen spalten bei Verfügbarkeit eines nächsten komplementären NT 3'-OH-Ester-Verbindungen, so dass die an die 3'-OH-Gruppe gebundene Markierung in die Lösung freigesetzt wird und nicht mehr als Terminator wirken kann (Rasolonjatovo et al. Nucleosides & Nucleotides 1999, v.18 S.1021, Canard et al. PNAS 1995 v.92 S.10859). In Positionen, an denen eine Polymerase mehrere gleiche NTs nacheinander einbauen kann, führt das zu einem fehlerhaften Signal. Im Abschlußbericht des BMBF-Verbundsprojekts „Sequenzierung mit Multiplexfarbstoffen und

Kapillarelektrophorese" G. Sagner, 1999, wurde berichtet, dass Modifikationen der 3'-Position von Nukleotiden zur Aufhebung ihrer Substrateigenschaften für Polymerasen geführt haben.

5 Die Schwierigkeiten bei der Entwicklung passender NT-Analoga für das Verfahren basieren auf folgenden Rahmenbedingungen:

- 1) Die Reaktion muss so gesteuert werden, dass die Polymerase NT*s einzeln einbaut (Stop des weiteren Einbaus).
- 2) NT*s müssen einen Farbstoff tragen, der den Anforderungen der 10 Detektion genügt.
- 3) Der Farbstoff muß unter milden Bedingungen abspaltbar sein, so dass weder die Genprodukt-Primer-Komplexe, noch einzelne Komponenten des Systems beschädigt werden.
- 4) Die Abspaltung muss möglichst schnell und quantitativ erfolgen.
- 5) Der Stop des Einbaus muss reversibel sein und unter milden 15 Bedingungen aufgehoben werden können.

Bis jetzt wurde keine praktisch brauchbare Lösung für diese Probleme in der einschlägigen Literatur vorgestellt.

20

Durch die vorliegende Erfindung werden nunmehr die im Stand der Technik bekannten Probleme erstmals gelöst. Erfindungsgemäß werden für die Sequenzierung NTs' mit einer sterisch anspruchsvollen Gruppe an der Base verwendet.

25

Eine an die Base gekoppelte sterisch anspruchsvolle Gruppe kann zur Behinderung der weiteren Synthese führen, wobei diese Behinderung in der Fachliteratur als unerwünschte Eigenschaft modifizierter NTs bei der Markierung von Nukleinsäuren 30 angesehen wurde. Biotin, Digoxigenin und Fluoreszenzfarbstoffe wie Fluoreszein, Tetramethylrhodamine und Cy3-Farbstoff stellen Beispiele einer solchen sterisch anspruchsvollen Gruppe dar (Zhu et al. Cytometry 1997, v.28, S.206, Zhu et al. NAR 1994, v.22, S.3418, Gebeyehu et al., NAR 1987, v.15, S.4513, Wiemann 35 et al. Analytical Biochemistry 1996, v.234, S.166, Heer et al. BioTechniques 1994 v.16 S.54).

Bei der Sequenzierungsreaktion im erfindungsgemäßen Verfahren werden markierte NTs' mit einer Polymerase und Nukleinsäureketten inkubiert. Die NTs' tragen dabei eine an die Base reversibel gekoppelte sterisch anspruchsvolle Gruppe. Wenn ein Reaktions-
5 gemisch, das nur modifizierte NTs' enthält, in der Reaktion eingesetzt wird, dann kann die Polymerase nur ein einziges NT' einbauen. Der Einbau eines nächsten NT' wird sterisch gehemmt. Diese NTs' treten somit als Terminatoren der Synthese auf. Nach der Entfernung der sterisch anspruchsvollen Gruppe kann das
10 nächste komplementäre NT' eingebaut werden. Weil diese NTs' kein absolutes Hindernis zur weiteren Synthese darstellen, sondern nur für den Einbau eines weiteren markierten NT', werden sie als Semiterminatoren bezeichnet.

15 4.4.2 Allgemeine Struktur des NT'

Ihre gemeinsamen Merkmale sind in Fig. 6a,b,d dargestellt. Diese Struktur ist dadurch charakterisiert, dass an der Base über einen spaltbaren Linker (A-E) eine sterische Gruppe (D)
20 und der Fluoreszenzmarker (F) gebunden sind.

Als Grundlage für die NTs' dienen Deoxynukleosid-Triphosphate mit Adenosin (A), Guanosin(G), Cytidin (C) und Uridin (U) als Nukleosidrest. Anstelle von Guanosin kann Inosin verwendet
25 werden.

4.4.3 Marker, Fluorophore

Jede Base ist mit einem für sie charakteristischen Marker (F) markiert (Fig. 6). Der Marker ist ein fluoreszierender Farbstoff. Mehrere Faktoren beeinflussen die Wahl des Fluoreszenzfarbstoffes. Die Wahl ist nicht eingeschränkt, sofern der Farbstoff folgenden Anforderungen genügt:

a) Die verwendete Detektionsapparatur muß diesen Marker als einziges Molekül gebunden an DNA unter milden Bedingungen (vorzugsweise Reaktionsbedingungen) identifizieren können. Die Farbstoffe haben vorzugsweise große Photostabilität. Ihre Fluoreszenz wird vorzugsweise von der DNA nicht oder nur unwesentlich gequenchet.

15

b) Der an das NT gebundene Farbstoff darf keine irreversible Störung der enzymatischen Reaktion verursachen.

c) mit dem Farbstoff markierte NTs' müssen von der Polymerase in die Nukleinsäurekette eingebaut werden.

20

d) Bei einer Markierung mit verschiedenen Farbstoffen sollen diese Farbstoffe keine beträchtlichen Überlappungen in ihren Emissionsspektren aufweisen.

25

Im Rahmen der vorliegenden Erfindung verwendbare Fluoreszenzfarbstoffe sind in "Handbook of Fluorescent Probes and Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes mit Strukturformeln zusammengestellt. Erfindungsgemäß werden vorzugsweise folgende Farbstoffklassen als Marker eingesetzt: Cyanin-Farbstoffe und deren Abkömmlinge (z.B. Cy2, Cy3, Cy5, Cy7 Amersham Pharmacia Biotech, Waggoner US-Patent 5.268.486), Rhodamine und deren Abkömmlinge (z.B. TAMRA, TRITC, RG6, R110, ROX, Molecular Probes), Xanthene-Derivate (z.B. Alexa 568, Alexa 594, Molecular Probes, Mao et al. US-Patent 6.130.101). Diese Farbstoffe sind kommerziell erhältlich.

35

Dabei kann man je nach spektralen Eigenschaften und vorhandener Apparatur entsprechende Farbstoffe auswählen. Die Farbstoffe werden an den Linker z.B. über Thiocyanat- oder Ester-Bindung gekoppelt ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Jameson et al. Methods in Enzymology 1997 v.278 S.363, Waggoner Methods in Enzymology 1995 v.246 S.362), s. Beispiele 1 und 2.

4.4.4 Natur der sterisch anspruchsvollen Gruppe

10

Die Gruppe (D) (Fig. 6a,b,d) stellt ein Hindernis für den Einbau eines weiteren komplementären markierten NT' durch eine Polymerase dar.

15 Biotin, Digoxigenin und Fluoreszenzfarbstoffe wie Fluoreszein, Tetramethylrhodamine und Cy3-Farbstoff stellen Beispiele einer solchen sterisch anspruchsvollen Gruppe dar (Zhu et al. Cytometry 1997, v.28, S.206, Zhu et al. NAR 1994, v.22, S.3418, Gebeyehu et al., NAR 1987, v.15, S.4513, Wiemann et al. 20 Analytical Biochemistry 1996, v.234, S.166, Heer et al. BioTechniques 1994 v.16 S.54). Die chemische Struktur dieser Gruppe ist nicht eingeschränkt, sofern sie den Einbau des markierten NT', an das sie geknüpft ist, nicht wesentlich stört und keine irreversible Störung der enzymatischen Reaktion 25 verursacht.

Diese Gruppe kann als selbständiger Teil im Linker (6a) auftreten oder mit dem Farbstoff (6b) oder der spaltbaren Gruppe (6d) identisch sein. Durch die Spaltung des Linkers wird diese 30 sterisch anspruchsvolle Gruppe (D) nach der Detektion des Signals entfernt, so dass die Polymerase ein weiteres markiertes NT' einbauen kann. Bei einer Struktur wie in 6d wird die sterische Gruppe durch die Spaltung beseitigt.

35 In einer bevorzugten Ausführungsform übernimmt der Fluoreszenzfarbstoff die Funktion einer solchen sterisch anspruchsvollen Gruppe, so dass ein markiertes Nukleotid eine in Fig. 6b,k,l

dargestellte Struktur aufweist.

In einer anderen bevorzugten Ausführungsform übernimmt die photolabile spaltbare Gruppe die Funktion einer solchen 5 sterisch anspruchsvollen Gruppe (Fig. 6d).

4.4.5 Linker

Der Marker (Fluoreszenzfarbstoff) ist an die Base vorzugsweise 10 über einen Abstandhalter unterschiedlicher Länge, einen sog. Linker, gebunden. Beispiele für Linker sind in Fig. 6e,f,g,h,i,j,k,l,m gegeben. Beispiele der Ankoppelung eines Linkers an die Base können aus folgenden Quellen entnommen werden (Hobbs et al. US Patent 5.047.519, Khan et al. US Patent 15 5.821.356, Klevan et al. US Patent 4.828.979, Hanna M. Method in Enzymology 1996 v.274, S.403, Zhu et al. NAR 1994 v.22 S.3418, Herman et al. Methods in Enzymology 1990 v.184 S.584, J.L.Ruth et al. Molecular Pharmacology 1981 v.20 S.415, L. Ötvös et al. NAR 1987 v.15 S.1763, G.E.Wright et al. Pharmac 20 Ther. 1990 v.47, S.447, „Nucleotide Analogs; Synthesis and Biological Function“ K.H. Scheit 1980, Wiley-Interscience Publication, "Nucleic acid chemistry" Ed. L.B.Townsend, v.1-4, Wiley-Interscience Publication, "Chemistry of Nucleosides and Nucleotides" Ed. L.B.Townsend, v.1-3, Plenum Press).

25

Die gesamte Länge des Linkers kann variieren. Sie entspricht der Anzahl der Kohlenstoff-Atome in den Abschnitten A, C, E (Fig. 6a,b,d) und liegt vorzugsweise zwischen 3 und 20. Optimalerweise beträgt sie zwischen 4 und 10 Atomen. Die 30 chemische Zusammensetzung des Linkers (Abschnitte A,C,E in Fig. 6a,b,d) ist nicht eingeschränkt, sofern sie unter Reaktionsbedingungen stabil bleibt und keine Störung der enzymatischen Reaktion verursacht.

35 4.4.6 Spaltbare Verbindung, Spaltung

Der Linker trägt eine spaltbare Verbindung oder spaltbare

Gruppe (Abschnitt (B) in Fig. 6a,b,d und Abschnitt (C) in Fig. 6k,l). Diese spaltbare Verbindung ermöglicht die Entfernung des Markers und des sterischen Hindernisses am Ende jedes Zyklus. Ihre Wahl ist nicht eingeschränkt, sofern sie unter den Bedingungen der enzymatischen Sequenzierungsreaktion stabil bleibt, keine irreversible Störung der Polymerase verursacht und unter milden Bedingungen abgespalten werden kann. Unter "milden Bedingungen" sind solche Bedingungen zu verstehen, die den Genprodukt-Primer-Komplex nicht zerstören, wobei z.B. der pH-Wert vorzugsweise zwischen 3 und 11 liegt, die Temperatur zwischen 0°C und einem Temperaturwert (x). Dieser Temperaturwert (x) hängt von der T_m des Genprodukt-Primer-Komplexes (T_m ist "melting Point") und wird beispielsweise als T_m (Genprodukt-Primer-Komplex) minus 5°C errechnet (z.B. T_m ist 47°C, dann liegt die maximale Temperatur bei 42°C; unter diesen Bedingungen eignen sich besonders Ester-, Thioester-, Disulfid-Verbindungen und photolabile Verbindungen als spaltbare Verbindungen).

Vorzugsweise gehört die genannte Verbindung zu chemisch oder enzymatisch spaltbaren oder photolabilen Verbindungen. Als Beispiele von chemisch spaltbaren Gruppen sind Ester-, Thioester- und Disulfid-Verbindungen bevorzugt (Fig. 6k,l) („Chemistry of protein conjugation and crosslinking" Shan S. Wong 1993 CRC Press Inc., Herman et al. Method in Enzymology 1990 v.184 S.584, Lomant et al. J.Mol.Biol. 1976 v.104 243, "Chemistry of carboxylic acid and esters" S.Patai 1969 Interscience Publ.). Beispiele für photolabile Verbindungen (Fig. 6m) können in folgenden Literaturstellen gefunden werden: "Protective groups in organic synthesis" 1991 John Willey & Sons, Inc., V. Pillai Synthesis 1980 S.1, V. Pillai Org.Photochem. 1987 v.9 S.225, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese" H. Giegrich, 1996, Konstanz, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese" S.M.Bühler, 1999, Konstanz).

Die Position der spaltbaren Verbindung/Gruppe im Linker ist

vorzugsweise nicht weiter als 10 Atome von der Base entfernt, noch bevorzugter nicht weiter als 3 Atome. Besonders bevorzugt liegt die spaltbare Verbindung oder Gruppe direkt an der Base.

- 5 Der Spaltungs- und Entfernungsschritt ist in jedem Zyklus vorhanden und muß unter milden Bedingungen (s.o.) verlaufen, so dass die Nukleinsäuren nicht beschädigt oder modifiziert werden.
- 10 Die Spaltung läuft bevorzugt chemisch (z.B. in milder saurer oder basischer Umgebung für eine Ester-Verbindung oder durch Zugabe eines Reduktionsmittels, z.B. Dithiothreitol oder Mercaptoethanol (Sigma) bei der Spaltung einer Disulfid-Verbindung), siehe Beispiel 1, oder physikalisch (z.B. durch
- 15 Beleuchtung der Oberfläche mit Licht einer bestimmten Wellenlänge für die Spaltung einer photolabilen Gruppe, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese“ H. Giegrich, 1996, Konstanz) ab.
- 20 Nach der Spaltung verbleibt an der Base ein Linkerrest (A) (Fig.6c). Falls die nach der Spaltung am Linkerrest freigeordnete Mercapto-Gruppe weitere Reaktionen stört, kann sie mit bekannten Mitteln chemisch modifiziert werden (wie z.B.
- 25 durch Disulfid- oder Iodacetatverbindungen).

Die Synthese eines spaltbaren Linkers wird an Beispielen gezeigt (vgl. Beispiele 1 und 2).

30 4.4.7 Kombination von Polymerase und NT'

- Insgesamt spielen die Größe, die Ladung und die chemische Struktur des Markers, die Länge des spaltbaren Linkers und des Linker-Rests sowie auch die Wahl der Polymerase eine wichtige
- 35 Rolle. Sie bestimmen gemeinsam, ob das markierte NT' durch die Polymerase in die wachsende Nukleinsäurekette eingebaut wird, und ob dadurch der Einbau des nächsten markierten NT' verhindert

wird. Zwei Bedingungen sind dabei besonders zu berücksichtigen:

Einerseits ist es wichtig, dass die Polymerase die Nukleinsäurekette mit dem eingebauten modifizierten NT' nach der
5 Spaltung des Linkers weiter verlängern kann. Es ist also wichtig, dass der Linkerrest "A" (Fig. 6c) nach der Spaltung keine wesentliche Störung für die weitere Synthese darstellt. Andererseits müssen eingebaute, nicht gespaltene NTs' ein Hindernis darstellen. Es können viele für die Reaktion
10 geeignete NTs' synthetisiert werden. Im einzelnen muß für jede Kombination aus Polymerase und NTs' eine Vorversuchsreihe durchgeführt werden, in der die Tauglichkeit eines bestimmten NT'-Typs für die Sequenzierung erprobt wird.

15 Die Pufferbedingungen werden nach Angaben des Polymeraseherstellers gewählt. Die Reaktionstemperatur wird für nicht thermostabile Polymerasen nach Angaben des Herstellers gewählt (z.B. 37°C für Sequenase Version 2), für thermostabile Polymerasen (z.B. Taq-Polymerase) beträgt die
20 Reaktionstemperatur maximal den Temperaturwert (x). Dieser Temperaturwert (x) hängt von der Tm des Genprodukt-Primer-Komplexes und wird z.B. als Tm (Genprodukt-Primer-Komplex) minus 5°C errechnet (z.B. Tm ist 47°C, dann liegt die maximale Reaktionstemperatur bei 42°C). Diese Pufferbedingungen und
25 Reaktionstemperatur werden weiter als "optimale Puffer- und Temperaturbedingungen" bezeichnet.

Die Reaktionszeit (entspricht der Dauer des Einbau-Schrittes in einem Zyklus) beträgt vorzugsweise weniger als eine Stunde,
30 idealerweise liegt die Reaktionszeit zwischen 10 sec und 10 min.

Als Beispiele von geeigneten Kombinationen zwischen NT' und Polymerase sind folgende Kombinationen zu nennen:

35

Falls DNA (z.B. cDNA) als Genprodukt in die Reaktion eingesetzt wird, können NT' mit einem kurzen Linkerrest (Synthese siehe

Beispiel 2, Fig. 6e,h,i): dNTP-SS-TRITC (L7), dNTP-SS-Cy3 (L11) und / oder NT' mit einem langen Linkerrest (Synthese siehe Beispiel 1, Fig. 6f,g,j): dNTP-SS-TRITC (L14) in Kombination mit Sequenase Version 2, Taq-Polymerase (GibcoBRL), ProHA-DNA-Polymerase (Eurogentec) oder Klenow Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech) eingesetzt werden.

Falls RNA (z.B. mRNA) als Genprodukt in die Reaktion eingesetzt wird, können NT' mit einem kurzen Linkerrest (Synthese siehe Beispiel 2, Fig. 6e,h,i): dNTP-SS-TRITC (L7), dNTP-SS-Cy3 (L11) und / oder NT' mit einem langen Linkerrest (Synthese siehe Beispiel 1, Fig. 6f,g,j): dNTP-SS-TRITC (L14) in Kombination mit AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse Transcriptase ohne RNase-Aktivität eingesetzt werden.

Die Tauglichkeit eines Linkerrests an der Base (A) für die Reaktion wird in einem Testsystem geprüft. Dabei werden gespaltene NTs' in eine Nukleinsäurekette nacheinander eingebaut. Man verwendet z.B. dUTP' mit dem gewünschten gespaltenen Linkerrest, poly-dA als Matrize für DNA-Polymerasen wie z.B. Sequenase Version 2, Taq-Polymerase, oder poly-A als Matrize für Reverse Transcriptasen wie z.B. AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), Oligo-dT20-Primer, die gewünschte Polymerase (s.o.) und führt unter für die jeweilige Polymerase geeigneten optimalen Puffer- und Temperaturbedingungen eine Reaktion durch. Die NT'-Konzentration liegt vorzugsweise zwischen 5µmol/l und 200µmol/l. Nach der Reaktion wird die Anzahl der in die Nukleinsäurekette eingebauten NTs' analysiert, z.B. durch die Auftrennung der Länge nach in einem Gel. Für die Rückschlüsse auf die Tauglichkeit des Linkerrests kann man folgende Angaben verwenden: Wenn die Polymerase mehr als 20 NTs' einbauen kann, so ist dieser Linkerrest für eine Sequenzierungsreaktion geeignet. Beim Einbau von weniger als 20 gespaltenen NTs' ist diese Kombination aus NT' und Polymerase nicht optimal für die

Sequenzierungsreaktion.

Wenn ein passender Linkerrest feststeht, wird in einem weiteren Testsystem geprüft, ob die markierten, nicht gespaltenen NTs' 5 als Semiterminatoren funktionieren. Das wird geprüft, indem die markierten NTs' unter für die Reaktion geeigneten optimalen Puffer- und Temperaturbedingungen mit der Polymerase und einer Matrize inkubiert werden. Die NT'-Konzentration liegt vorzugsweise zwischen 5 $\mu\text{mol/l}$ und 200 $\mu\text{mol/l}$. Die Matrize ist so zu 10 wählen, dass der Einbau mehrerer NTs' nacheinander zu erwarten wäre, z.B. für dUTP' kann man polydA oder poly-A, wie im oben dargestellten Beispiel verwenden. Idealerweise baut die Polymerase nur ein einziges NT' ein.

15 Falls bei gegebenen optimalen Puffer- und Temperaturbedingungen durch eine Polymerase mehrere NTs' nacheinander eingebaut werden, kann man die Reaktionsparameter (z.B. NT'-Konzentration, Reaktionstemperatur) verändern und der jeweiligen Kombination aus Polymerase und NT' anpassen. Das wichtigste dabei ist, dass 20 die Polymerase in der vorgegebenen Zeit (liegt vorzugsweise zwischen 10 sec und 10 min) ein zweites NT' nicht einbaut. Erfindungsgemäß erfolgt diese Anpassung in einer Ausführungsform durch die Veränderung der Reaktionstemperatur. Die anderen Parameter der Reaktion werden dabei konstant 25 gehalten.

Falls DNA als Genprodukt (z.B. cDNA) in die Sequenzierungsreaktion eingesetzt wird, können als Polymerasen beispielsweise Polymerasen verwendet werden, die DNA-Polymerasen ohne 3'-5'- 30 Exonukleaseaktivität sind, vorzugsweise aus der Gruppe bestehend aus viralen DNA-Polymerasen vom Sequenase-Typ, bakteriellen DNA-Polymerasen I, thermostabilen DNA-Polymerasen und deren Modifikationen. Insbesondere kommen Sequenase Version 2, Klenow Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' 35 Exonukleaseaktivität, Taq-Polymerase oder ProHA-DNA-Polymerase in Betracht.

Falls RNA als Genprodukt (z.B. mRNA) in die Sequenzierungs-

reaktion eingesetzt wird, können handelsübliche RNA-abhängige DNA-Polymerasen eingesetzt werden, z.B. AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse Transcriptase ohne RNase-Aktivität.

5

Die NT'-Konzentration liegt bei diesen Experimenten üblicherweise zwischen $5\mu\text{mol/l}$ und $200\mu\text{mol/l}$, vorzugsweise zwischen $10\mu\text{mol/l}$ und $100\mu\text{mol/l}$. Die Konzentration der Polymerase und die Pufferbedingungen werden nach Angaben vom Hersteller
10 gewählt. Die Dauer der Reaktion kann variieren und liegt vorzugsweise zwischen 10 sec und 10 min, was der Dauer des Einbau-Schrittes (b) in einem Zyklus entsprechen würde. Bei nicht thermostabilen Polymerasen wie z.B. Sequenase Version 2 (Amersham Pharmacia Biotech), Klenow-Fragment der DNA Polymerase I ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia
15 se I ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech), AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse Transcriptase ohne RNase-Aktivität wird die Reaktionstemperatur von konventionellen 37°C vorzugsweise auf 20°C bis 30°C reduziert. Bei thermostabilen
20 Polymerasen wie z.B. Taq-Polymerase (GibcoBRL), ProHA-DNA-Polymerase (Eurogentec) wird die Reaktionstemperatur von konventionellen $70-75^\circ\text{C}$ vorzugsweise auf Werte reduziert, die zwischen 30°C und dem Temperaturwert (x) liegen. Dieser Temperaturwert (x) hängt von der T_m des Genprodukt-Primer-Komplexes und wird als T_m (Genprodukt-Primer-Komplex) minus 5°C
25 errechnet (z.B. T_m ist 47°C , dann liegt der Temperaturwert (x) bei 42°C).

In einer anderen bevorzugten Ausführungsform der Erfindung erfolgt die Anpassung der Reaktionsbedingungen durch die
30 Verminderung der NT'-Konzentration auf unter $5\mu\text{mol/l}$, die anderen Parameter der Reaktion (Pufferbedingungen, Temperaturbedingungen) werden konstant gehalten. Die Konzentration der NT's liegt vorzugsweise bei dieser Anpassung
35 zwischen 0.5 und $5\mu\text{mol/l}$. Die Dauer der Reaktion liegt zwischen 10 sec und 10 min. Das wichtigste bei der Wahl der NT'-

Konzentration ist, dass die Polymerase in der vorgegebenen Zeit (liegt vorzugsweise zwischen 10 sec und 10 min) ein zweites NT^{*} nicht einbaut.

- 5 Nach Optimierung der Reaktionsbedingungen für den Einbau eines einzelnen NT^{*} muß man die Reaktion mit gespaltenen NTs^{*} wiederholen. Unter entsprechend geänderten Reaktionsparametern muß die Polymerase die gespaltenen NTs^{*} nacheinander einbauen können.

10

Die Optimierungsreaktion korreliert mit dem Einbauschnitt, Schritt (b), in einem Zyklus. Die für die Optimierungsreaktion ermittelten Bedingungen die Temperatur, die Konzentration an NT^{*}, die Pufferbedingungen und die Dauer der Reaktion werden für 15 die Reaktion auf der Oberfläche übernommen.

Unter diesen Reaktionsbedingungen erfolgt der Einbau von NT^{*} in die zu Genprodukten komplementären Stränge vorzugsweise so, dass an mehr als 50% der an der Sequenzierungsreaktion 20 beteiligten Genprodukte (extensionsfähige Genprodukt-Primer-Komplexe) in einem Zyklus ein markiertes NT^{*} eingebaut wird, vorzugsweise an mehr als 90%. Das hängt damit zusammen, dass an manchen Nukleinsäureketten die Reaktion sehr langsam abläuft. Ein Einbau der NTs^{*} an jeder komplementären Position in jedem 25 Zyklus wird angestrebt, ist aber nicht erforderlich, weil nur die erfolgreichen Einbaureaktionen detektiert und ausgewertet werden; eine verzögerte Reaktion im Nachfolgenden Zyklus führt nicht zu einem Sequenzierungsfehler.

- 30 Vorzugsweise wird für alle NTs^{*} dieselbe Polymerase verwendet. Es können aber auch verschiedene Polymerasen für verschiedene NTs^{*} eingesetzt werden.

4.4.8 Farbiges Kodierungsschema, Anzahl der Farbstoffe

Einen Zyklus kann man durchführen mit:

- 5 a) vier verschieden markierten NT*s
- b) zwei verschieden markierten NT*s
- c) einem markierten NT*
- d) zwei verschieden markierten NT*s und zwei unmarkierten NTs,

10

d.h.

a) Man kann alle 4 NTs mit verschiedenen Farbstoffen markieren und alle 4 gleichzeitig in die Reaktion einsetzen. Dabei erreicht man die Sequenzierung einer Nukleinsäurekette mit einer minimalen Anzahl von Zyklen. Diese Variante der Erfindung stellt allerdings hohe Anforderungen an das Detektionssystem: 4 verschiedene Farbstoffe müssen in jedem Zyklus identifiziert werden.

20

b) Zur Vereinfachung der Detektion kann eine Markierung mit zwei Farbstoffen gewählt werden. Dabei werden 2 Paare von NTs* gebildet, die jeweils verschieden markiert sind, z.B. A und G tragen die Markierung "X", C und U tragen die Markierung "Y". In die Reaktion in einem Zyklus (n) werden 2 unterschiedlich markierte NTs* gleichzeitig eingesetzt, z.B. C* in Kombination mit A*, und im darauffolgenden Zyklus (n+1) werden dann U* und G* zugegeben.

30 c) Man kann auch nur einen einzigen Farbstoff zur Markierung aller 4 NTs* verwenden und pro Zyklus nur ein NT* einsetzen.

d) In einer technisch vereinfachten Ausführungsform werden 35 pro Zyklus zwei unterschiedlich markierte NT*s eingesetzt und zwei unmarkierte NTs (sogen. 2NT*s / 2NTs-Methode).

4.5 Detektionsapparatur

Einzelne Moleküle auf einer Oberfläche kann man mit verschiedenen Methoden untersuchen. Es sind mehrere Verfahren bekannt: z.B. AtomForce-Mikroskopie, Elektronen-Mikroskopie, Nahfeld-Fluoreszenz-Mikroskopie, Weitfeld-Fluoreszenz-Mikroskopie, TIR-Mikroskopie usw. (Science 1999 v.283 1667, Unger et al. BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, 10 Dickson et al. Science 1996 v.274 S.966, Xie et al. Science 1994 v.265 S.361, Nie et al. Science 1994 v.266 S.1018, Betzig et al. Science 1993 v.262 S.1422).

Erfindungsgemäß werden Fluoreszenz-Signale einzelner in die 15 Nukleinsäurekette eingebauter NTs' vorzugsweise mit einem Weitfeld-Fluoreszenz-Mikroskop (Epifluoreszenz) oder einem Laser-Scanning-Mikroskop oder einem TIRF-Mikroskop (Total Internal Reflection Fluorescence Microscope) (Weitfeld-Optik-Detektionssystem).

20

Es sind verschiedene Varianten der Konstruktion einer solchen Apparatur möglich (Weston et al. J.Chem.Phys. 1998 v.109 S.7474, Trabesinger et al. Anal. Chem. 1999 v.71 S.279, Adachi et al. Journal of Microscopy 1999 v.195 S.125, Unger et al. 25 BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Tokunaga et al. Bichem.Biophys.Res.Com. 1997 v.235 S.47, "Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspec- 30 troscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Unterschiede in ihrem konkreten Aufbau ergeben sich aus der Variation ihrer Einzelteile. Die Vorrichtung für 35 das Anregungslicht kann z.B. auf der Basis eines Lasers, einer Lampe oder von Dioden funktionieren. Für die Detektionsvorrichtung können sowohl CCD-Kameras als auch PMT dienen. Andere

Beispiele für technische Details siehe ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence
5 microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Es ist nicht die Aufgabe dieser Erfindung, alle möglichen technischen Varianten einer Detektionsvorrichtung aufzuzählen. Der prinzipielle Aufbau einer geeigneten Apparatur
10 wird in einem Schema Fig. 7 erläutert. Sie besteht aus folgenden Elementen:

- Lichtquelle zur Anregung der Fluoreszenz (1)
 - Lichtleitender Teil (2)
 - 15 Scantisch (3)
 - Vorrichtung zur Selektion von Spektren (4)
 - Detektionsvorrichtung (5)
 - Computer mit Steuerungs- und Analysefunktionen (6)
- 20 Diese Elemente der Apparatur können kommerziell erworben werden (Mikroskop-Firmen: Zeiss, Leica, Nikon, Olympus).

Im folgenden soll beispielsweise eine für die Detektion einzelner Moleküle geeignete Kombination aus diesen Elementen vor-
25 gestellt werden:

- Weitfeld-Fluoreszenz-Mikroskop Axioplan 2 (Zeiss) mit Quecksilberdampflampe
- Objektiv Planneofluar 100x, NA 1.4 (Zeiss)
- 30 Kamera SenSys™ (Photometrix) oder AxioCam (Zeiss)
- Computer mit Software zur Steuerung und Analyse

Nachfolgend soll die Vorgehensweise bei der Detektion erläutert werden. Man beachte dabei die allgemeinen Regeln der Fluoreszenzmikroskopie ("Confocal Laser Scanning Microscopy" 1997
35 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC

Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press).

5 Die Detektion umfaßt folgende Phasen:

- 1) Vorbereitung zur Detektion
- 2) Durchführung eines Detektionsschrittes in jedem Zyklus, wobei jeder Detektionsschritt als Scanvorgang abläuft und
10 folgende Operationen umfaßt:
 - a) Einstellung der Position des Objektivs (X,Y-Achse),
 - b) Einstellung der Fokusebene (Z-Achse),
 - c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum jeweiligen
15 Genprodukt,
 - d) Verschiebung zur nächsten Position auf der Oberfläche.

Die Signale von in die den Genprodukten komplementären Stränge eingebauten NTs' werden durch das Abscannen der Oberfläche
20 registriert. Der Scanvorgang kann in verschiedener Weise ausgeführt werden ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS
25 Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Beispielsweise wird ein diskontinuierlicher Scanvorgang gewählt. Dabei wird das Objektiv schrittweise über die Oberfläche bewegt (Fig. 8a), so dass von jeder Oberflächenposition ein zweidimensionales Bild
30 (2D-Bild) entsteht (Fig. 8b,c), für Versuchsanordnung s. Beispiel 5 .

Dieses 2D-Bild kann mit verschiedenen Methoden erstellt werden: z.B. durch den Laser-Scan einer Position des Mikroskopfeldes
35 (Laser-Scanning-Mikroskopie) oder durch eine Kameraaufnahme an einer Position (vgl. Handbücher der Mikroskopie). Als Beispiel wird die Detektion einzelner Moleküle mit einer CCD-Kamera

beschrieben.

1) Vorbereitung zur Detektion:

Am Anfang wird festgelegt, wie viele Kopien der Genprodukte zur
5 Expressionsanalyse notwendig sind. Mehrere Faktoren spielen
dabei eine Rolle. Die genaue Zahl hängt z.B. von der relativen
Präsenz der Genprodukte im Ansatz und von der gewünschten
Genauigkeit der Analyse ab. Die Anzahl der analysierten Gen-
produkte liegt vorzugsweise zwischen 1000 und 10.000.000. Für
10 stark exprimierte Gene kann die Anzahl der analysierten Gen-
produkte niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse
schwach exprimierter Gene muß sie erhöht werden, z.B. auf
100.000 oder noch weiter.

Es werden beispielsweise 100.000 einzelne Genprodukte gleichzei-
15 tig analysiert. Dabei werden auch schwach exprimierte Gene (mit
z.B. ca.100 mRNA-Molekülen/Zelle, was ca. 0.02% Gesamt-mRNA
entspricht) in der Reaktion mit durchschnittlich 20 identifi-
zierten Genprodukten repräsentiert.

20 2) Durchführung eines Detektionsschrittes in jedem Zyklus

Zur Sequenzierung müssen die Positionen der Genprodukt-Primer-
Komplexe bestimmt werden, damit man eine Grundlage für die
Zuordnung der Signale hat. Die Kenntnis dieser Positionen
erlaubt eine Aussage darüber, ob die Signale einzelner Moleküle
25 von eingebauten NTs' stammen oder von zufällig an die Oberfläche
gebundenen NTs'. Diese Positionen können mit verschiedenen
Methoden identifiziert werden.

In einer bevorzugten Ausführungsform werden die Positionen
30 immobilisierter Genprodukte während der Sequenzierung
identifiziert. Dabei wird die Tatsache genutzt, dass die
Signale von den in die Nukleinsäurekette eingebauten NTs' immer
dieselben Koordinaten haben. Das ist durch die Fixierung der
Nukleinsäureketten gewährleistet. Die unspezifisch gebundenen
35 NTs' binden zufällig an verschiedenen Stellen der Oberfläche.

Zur Identifizierung der Positionen von fixierten Genprodukten

werden die Signale auf Übereinstimmung ihrer Koordinaten aus mehreren aufeinander folgenden Zyklen überprüft. Das kann z.B. am Anfang der Sequenzierung erfolgen. Die übereinstimmende Koordinaten werden als Koordinaten der Genprodukte bewertet und
5 gespeichert.

Das Scan-System muß reproduzierbar über mehrere Zyklen die Oberfläche abscannen können. X,Y und Z-Achsen-Einstellungen an jeder Oberflächenposition können von einem Computer
10 kontrolliert werden. Stabilität und Reproduzierbarkeit der Einstellung von Objektivpositionen in jedem Scanvorgang entscheiden über die Qualität der Detektion und somit über die Identifizierung der Signale einzelner Moleküle.

15 a) Einstellung der Position des Objektivs (X,Y-Achse)

Die mechanische Instabilität der kommerziell erhältlichen Scantische und die geringe Reproduzierbarkeit der wiederholten Einstellung derselben X,Y-Positionen machen eine präzise Analysen
20 der Signale einzelner Moleküle über mehrere Zyklen schwierig. Es existieren viele Möglichkeiten, eine Übereinstimmung der Koordinaten bei wiederholten Einstellungen zu verbessern bzw. mögliche Abweichungen zu kontrollieren. Als Beispiel wird eine Kontrollmöglichkeit angeführt. Nach einer groben mechanischen
25 Einstellung der Objektivposition wird ein Kontrollbild von einem mit der Oberfläche fest verbundenen Muster gemacht. Auch wenn die mechanische Einstellung nicht exakt dieselben Koordinaten aufweist (Abweichungen bis zu 10 µm sind durchaus möglich), kann man mittels optischer Kontrolle eine Korrektur
30 vornehmen. Das Kontrollbild vom Muster dient als Koordinatensystem für das Bild mit Signalen von eingebauten NTs'. Eine Voraussetzung für eine solche Korrektur ist, dass keine weiteren Bewegungen der Oberfläche zwischen diesen beiden Aufnahmen gemacht werden. Signale von einzelnen Molekülen
35 werden in Relation zum Muster gesetzt, so dass eine X,Y-Abweichung in der Musterposition gleiche X,Y-Abweichung in der Position der Signale einzelner Moleküle bedeutet. Das

Kontrollbild vom Muster kann vor, während oder nach der Detektion einzelner Moleküle gemacht werden. Ein solches Kontrollbild muß entsprechend bei jeder Einstellung auf einer neuen Oberflächenposition gemacht werden.

5

b) Einstellung der Fokusebene (Z-Achse)

Die Oberfläche ist nicht absolut flach und weist verschiedene Unebenheiten auf. Dadurch verändert sich der Oberflächen-Objektiv-Abstand beim abscannen benachbarter Stellen. Diese Unterschiede im Abstand können dazu führen, dass einzelne Moleküle die Fokusebene verlassen und so der Detektion entgehen.

Aus diesem Grund ist es wichtig, dass beim Abscannen der Oberfläche eine reproduzierbare Einstellung der Fokusebene an jeder Objektivposition erreicht wird.

Es gibt verschiedene Möglichkeiten, die Fokusebene reproduzierbar einzustellen. Beispielsweise kann folgende Methode angewendet werden: Da die Anregung einzelner Moleküle zum Auslösen ihrer Fluoreszenz führen kann, wird auf die Oberfläche ein Marker aufgebracht, der zur Einstellung der Fokusebene dient. Danach erfolgt die Detektion der Signale einzelner Moleküle. Der Marker kann beliebiger Natur sein (z.B. Farbstoff oder Muster), darf aber die Detektion und die Reaktion nicht beeinträchtigen.

c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum jeweiligen Genprodukt.

Das mit Hilfe des Detektionssystems erzeugte zweidimensionale Bild der Reaktionsoberfläche enthält die Signalinformationen von in die Genprodukte eingebauten NT's. Diese müssen vor der weiteren Verarbeitung aus der Gesamtdatenmenge der Bildinformationen mit geeigneten Methoden extrahiert werden. Die dazu notwendigen Algorithmen zur Skalierung, Transformation und

Filterung der Bildinformationen zählen zum Standardrepertoire der digitalen Bildverarbeitung und Mustererkennung (Haberäcker P. "Praxis der Digitalen Bildverarbeitung und Mustererkennung". Hanser-Verlag, München, Wien, 1995; Galbiati L.J. "Machine vision and digital image processing fundamentals". Prentice Hall, Englewood Cliffs, New Jersey, 1990). Die Signalextraktion erfolgt vorzugsweise über ein Grauwertbild, das die Helligkeitsverteilung der Reaktionsoberfläche für den jeweiligen Fluoreszenzkanal abbildet. Wenn bei der Sequenzierungsreaktion mehrere Nukleotide mit unterschiedlichen Fluoreszenz-Farbstoffen verwendet werden, kann zunächst für jedes verwendete fluorezenzmarkierte Nukleotid (A,T,C,G oder U) ein separates Grauwert-Bild erzeugt werden. Dafür können prinzipiell 2 Verfahren angewendet werden:

15

1. Durch Verwendung von geeigneten Filtern (Zeiss-Filtersätze) wird für jeden Fluoreszenzkanal ein Grauwertbild erzeugt.

2. Aus einem aufgenommenen Mehrkanal-Farb-Bild werden mit Hilfe eines geeigneten Algorithmus durch ein Bildverarbeitungsprogramm die relevanten Farbkanäle extrahiert und jeweils als Grauwertbild einzeln weiterverarbeitet. Zur Kanalextraktion wird dabei ein für den jeweiligen Kanal spezifischer Farb-Schwellwertalgorithmus eingesetzt. So entstehen zunächst aus einem Mehrkanal-Farbbild einzelne Grauwertbilder 1 bis N. Diese Bilder definieren sich wie folgt:

$GB_N = (s(x,y))$ einkanalisches Grauwertbild
 $N = \{1, \dots, \text{Anzahl der Fluoreszenzkanäle}\}$.
 30 $M = \{0, 1, \dots, 255\}$ Grauwertmenge
 $S = (s(x,y))$ Bildmatrix des Grauwertbildes
 $x = 0, 1, \dots, L-1$ Bildzeilen
 $y = 0, 1, \dots, R-1$ Bildspalten
 (x,y) Ortskoordinaten eines Bildpunktes
 35 $s(x,y) \in M$ Grauwert des Bildpunktes.

Aus dieser Datenmenge wird nun durch ein geeignetes Programm

die relevante Bildinformation extrahiert. Ein solches Programm sollte folgende Arbeitsschritte realisieren:

Für GB_1 bis GB_n durchführen:

5

I. Vorverarbeitung des Bildes, so zum Beispiel gegebenenfalls Reduktion des durch die Digitalisierung der Bildinformation entstandenen Bildrauschens, etwa durch Grauwertglättung.

- 10 II. Prüfung jedes Bildpunkt (x,y) des Grauwertbildes, ob dieser Punkt im Zusammenhang mit den ihn umgebenden unmittelbaren und weiter entfernten Nachbarbildpunkten die Eigenschaften eines Fluoreszenzpunktes erfüllt. Diese Eigenschaften hängen unter anderem von der verwendeten
15 Detektionsapparatur und der Auflösung des Grauwertbildes ab. Sie können beispielsweise ein typisches Verteilungsmuster von Helligkeits-Intensitätswerten über einer den Bildpunkt umgebenden Matrix darstellen. Die dazu verwendbaren Methoden der Bildsegmentierung reichen von einfachen Schwellwert-
20 verfahren bis hin zur Verwendung neuronaler Netze.

Erfüllt ein Bildpunkt (x,y) diese Anforderungen, dann folgt ein Vergleich mit den Koordinaten von in bisher durchgeführten Sequenzierungszyklen identifizierten Genprodukte. Bei einer
25 Übereinstimmung erfolgt die Zuordnung des Signals mit dem aus dem jeweiligen Fluoreszenzkanal hervorgehenden Nukleotid zu diesem Genprodukt. Signale mit nicht übereinstimmenden Koordinaten werden als Hintergrundsignale bewertet und verworfen. Die Analyse der Signale kann parallel zum Scanvorgang erfolgen.

30

In einer beispielhaften Ausführung wurde ein acht-Bit-Grauwertbild mit einer Auflösung von 1317 x 1035 Pixel verwendet. Um die durch die Digitalisierung entstandenen Veränderungen am Bild zu reduzieren, erfolgte zunächst eine Vorverarbeitung des
35 Gesamtbildes: Jedem Bildpunkt wurde der Mittelwert der Helligkeiten seiner 8-Nachbarn zugewiesen. Bei der gewählten Auflösung entsteht dadurch ein für einen Fluoreszenzpunkt

typisches Muster eines zentralen Bildpunkt mit dem größten Helligkeitswert und Nachbarbildpunkten mit nach allen Seiten hin abfallenden Helligkeiten. Erfüllte ein Bildpunkt diese Kriterien und überschritt der zentrifugale Helligkeitsabfall einen bestimmten Schwellenwert (zur Exklusion zu schwacher Fluoreszenzpunkte), dann wurde dieser zentrale Bildpunkt als Koordinate eines Fluoreszenzpunktes gewertet.

d) Verschiebung des Objektivs zur nächsten Position auf der Oberfläche. Nach der Detektion der Signale einzelner Moleküle wird das Objektiv über einer anderen Position der Oberfläche positioniert.

Insgesamt kann beispielsweise eine Folge von Aufnahmen mit der Kontrolle der X,Y-Position, der Einstellung der Fokusebene und mit der Detektion einzelner Moleküle bei jeder neuen Objektivposition gemacht werden. Diese Schritte können durch einen Computer gesteuert werden.

4.6 Zeitlicher Ablauf der Verfahrensschritte

Der Scanvorgang sowie die biochemische Reaktion nehmen eine gewisse Zeit in Anspruch. Wenn man diese Vorgänge nacheinander schaltet, kann man eine optimale Leistung der Apparatur erreichen. In einer bevorzugten Ausführung wird die Reaktion auf zwei getrennten Oberflächen durchgeführt (Fig. 9).

Als Beispiel kann eine Oberfläche mit gebundenen Genprodukten in 2 räumlich isolierte Teile getrennt werden, so dass Reaktionen auf diesen beiden Teilen unabhängig voneinander ablaufen können. In einem anderen Beispiel können Genprodukte auch von vornherein auf 2 getrennten Oberflächen immobilisiert werden.

Danach wird die Reaktion gestartet. Das Prinzip dabei ist, dass während auf einem Teil der Oberfläche die Reaktions- und Waschschritte ablaufen, der zweite Teil abgescannt wird. Dadurch

kann man einen kontinuierlichen Ablauf der Analyse erreichen und die Geschwindigkeit der Sequenzierung steigern.

Die Anzahl der Oberflächen, auf denen die Reaktion abläuft, kann auch größer als 2 sein. Das erscheint dann sinnvoll, wenn die Reaktion als zeitlich limitierender Schritt auftritt, d.h. die Detektion der Signale auf der Oberfläche schneller als die Reaktions- und Waschschrte abläuft. Um die Gesamtdauer der Reaktion an die Detektionsdauer anzupassen, kann jeder einzelne Schritt der Reaktion auf einer einzelnen Oberfläche mit einer zeitlichen Verzögerung im Vergleich zur nächsten Oberfläche ablaufen.

15 4.7 Analyse

Die gewonnenen Daten (kurze Sequenzen) werden mit Hilfe eines Programms mit bekannten Gensequenzen verglichen. Einem solchen Programm kann z.B. ein BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Durch die Wahl der Methode zur Materialvorbereitung wird unter anderem festgelegt, in welchen Abschnitten der Genprodukte die Sequenzen ermittelt werden und zu welchem Strang (sense oder antisense) sie gehören. Z.B. werden bei der Verwendung der polyA-Strecken als Primerbindungsstelle in mRNA Sequenzen aus NTRs (non-translating-regions) bestimmt. Bei der Verwendung der Methode mit antisense-cDNA als Matrize stammen die ermittelten Sequenzen unter anderem aus den proteinkodierenden Bereichen der Genprodukte.

Bei einer bevorzugten einfachen Variante der Erfindung wird die Genexpression nur qualitativ bestimmt. Dabei ist nur die Tatsache der Expression bestimmter Gene von Bedeutung.

Bei einer anderen bevorzugten Ausführungsform ist eine

quantitative Bestimmung der Verhältnisse zwischen einzelnen Genprodukten im Ansatz von Interesse. Es ist bekannt, dass die Aktivität eines Gens in einer Zelle durch eine Population identischer mRNA-Moleküle repräsentiert ist. In einer Zelle sind viele Gene gleichzeitig aktiv und werden dabei unterschiedlich stark exprimiert, was zum Vorhandensein vieler verschiedener unterschiedlich stark repräsentierter mRNA-Populationen führt.

10 Im folgenden wird auf die quantitative Analyse der Genexpression näher eingegangen:

Für eine quantitative Analyse der Genexpression werden die Abundanzen (Häufigkeiten) einzelner Genprodukte in der Sequenzierungsreaktion bestimmt. Dabei sind die Produkte stark exprimierter Gene in der Sequenzierungsreaktion häufiger vertreten als die schwach exprimierter Gene.

Nach der Zuordnung der Sequenzen zu bestimmten Genen wird der Anteil bzw. die Anzahl der ermittelten Sequenzen für jedes einzelne Gen bzw. Genprodukt bestimmt. Gene mit starker Expression haben einen höheren Anteil an der Gesamtpopulation der Genprodukte als Gene mit schwacher Expression.

25 Die Anzahl der analysierten Genprodukte liegt vorzugsweise zwischen 1000 und 10.000.000. Die genaue Anzahl der zu analysierenden Genprodukte hängt von der Aufgabenstellung ab. Für stark exprimierte Gene kann sie niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse schwach exprimierter Gene muß sie erhöht werden, z.B. auf 100.000 oder höher.

Werden beispielsweise 100.000 einzelne Genprodukte gleichzeitig analysiert, sind auch schwach exprimierte Gene, wie z.B. ca. 100 mRNA-Moleküle/Zelle (was ca. 0.02% Gesamt-mRNA entspricht), in der Reaktion mit durchschnittlich 20 identifizierten Genprodukten repräsentiert.

Als interne Kontrolle der Hybridisierung, der Immobilisation und der Sequenzierungsreaktion läßt sich folgende Methode verwenden:

5 Es können eine oder mehrere Nukleinsäureketten mit bekannten Sequenzen als Kontrolle eingesetzt werden (quantitativer Marker). Die Zusammensetzung dieser Kontrollsequenzen ist nicht eingeschränkt, sofern sie die Identifizierung der Genprodukte nicht stören. Bei der Sequenzanalyse der mRNA-Proben werden
10 RNA-Kontrollproben, bei der Analyse der cDNA-Proben entsprechend DNA-Kontrollproben eingesetzt. Diese Proben werden vorzugsweise bei allen Schritten mitgeführt. Sie können z.B. nach der mRNA-Isolation zugegeben werden. Im allgemeinen werden die Kontrollproben in gleicher Weise zur Sequenzanalyse
15 vorbereitet wie die zu analysierenden Genprodukte.

Die Kontrollsequenzen werden in bekannten, fest eingestellten (vorbestimmten) Konzentrationen zu den zu analysierenden Genprodukten zugegeben. Konzentrationen der Kontrollproben können
20 unterschiedlich sein, vorzugsweise liegen diese Konzentrationen (d.h. die Gesamtkonzentration der Kontrollsequenzen) zwischen 0.01% und 10% der Gesamtkonzentration der zu analysierenden Probe (100%). Beträgt die Konzentration der mRNA beispielsweise 10ng/µl, dann liegen die Konzentrationen von Kontrollproben
25 zwischen 1pg/µl und 1ng/µl.

Bei der quantitativen Analyse der Genexpression muß auch die allgemeine metabolische Aktivität der Zellen berücksichtigt werden, insbesondere, wenn ein Vergleich der Expression
30 bestimmter Gene bei verschiedenen äußeren Bedingungen angestrebt wird.

Die Veränderung im Expressionsniveau eines bestimmten Gens kann als Folge der Veränderung in der Transkriptionsrate dieses Gens
35 oder als Folge einer globalen Veränderung der Genexpression in der Zelle auftreten. Zur Beobachtung der metabolischen Zustände in der Zelle kann man die Expression der sogenannten "House-

keeping-Gene" analysieren. Beim Mangel an wichtigen Metaboliten ist beispielsweise das allgemeine Expressionsniveau in der Zelle niedrig, so dass auch konstitutiv exprimierte Gene ein niedriges Expressionsniveau haben.

5

Im Prinzip können alle konstitutiv exprimierten Gene als "House-keeping-Gene" dienen. Als Beispiele seien das Transferrin-Rezeptor-Gen oder das Beta-Aktin-Gen genannt.

Die Expression dieser House-keeping-Gene dient somit als
10 Bezugsgröße für die Analyse der Expression anderer Gene. Die Sequenzermittlung und Quantifizierung der Expression der House-keeping-Gene ist vorzugsweise ein Bestandteil des Analyse-Programms für die Genexpression.

15 Die Erfindung wird nachfolgend anhand von Beispielen verdeutlicht.

Beispiele

20

Beispiel 1:

Modifiziertes dUTP mit einem langen spaltbaren Linker (Fig. 6f-1) Als Ausgangssubstanzen dienen 5-(3-Aminoallyl)-2'-deoxyuridin-5'-triphosphat, AA-dUTP, (Sigma), 3,3'-Dithio-bis(propionsäure-N-Hydroxysuccinimidester), DTBP-NHS, (Sigma), 2-Mercaptoethylamin, MEA, (Sigma). Zu 100 µl 50mmol/l Lösung von AA-dUTP in 100mmol/l Borat-Puffer pH 8.5 werden 3 Äquivalente an DTBP-NHS in DMF (25 µl 0.4mol/l Lösung) zugegeben. Das Reaktionsgemisch wird 4 Stunden bei RT. inkubiert. Anschließend wird konz. Ammoniumacetat-Lösung (pH 9) zugegeben bis die Gesamtkonzentration an $\text{CH}_3\text{COONH}_4$ in der Reaktionslösung 100mmol/l ist, und die Reaktion wird eine weitere Stunde inkubiert. Danach werden zu diesem Gemisch 200 µl 1mol/l MEA-Lösung, pH 9, zugegeben und eine Stunde bei RT inkubiert. Anschließend wird zu diesem Gemisch solange eine gesättigte Lösung an I_2 in 0.3M KI-Lösung zugetropft, bis die Jodfarbe

bestehen bleibt. Die modifizierten Nukleotide werden auf einer DEAE-Cellulose-Säule in Ammoniumcarbonat-Gradient (pH 8.5) von anderen Reaktionsprodukten abgetrennt. Isolierung des Nukleotids mit dem spaltbaren Linker erfolgt auf RP-HPLC. An diesen Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Waggoner Method in Enzymology 1995 v.246, S.362, Jameson et al. Method in Enzymology 1997, v.278, S.363).

10

Auch andere Nukleotidanaloga (z.B. nach Hobbs et al. US Patent 5,047,519, Khan et al. US Patent 5,821,356) können in die Reaktion eingesetzt werden, so dass Nukleotidanaloga mit Strukturen in Fig. 6f-2,3,4 und 6g-1,2 erzeugt werden können.

15

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung von TRITC (Tetramethylrhodamin-5-isothiocyanat, Molecular Probes) angegeben (NT'-Struktur Fig. 6j):

20 Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 30 µl 100mmol/l Natrium-Borat-Puffer, pH 9, aufgelöst (10mmol/l NT'). Dazu werden 10 µl 10mmol/l TRITC in DMF gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT' erfolgt über RP-HPLC in Methanol-Wasser
25 Gradient. In ähnlicher Weise können andere Farbstoffe an die Aminogruppe des Linkers gekoppelt werden.

Das so hergestellte NT* erfüllt die Anforderungen des Einbaus in den DNA-Strang, des Fluoreszenznachweises und
30 Kettenabbruchs nach dem Einbau und der Aufhebung der Hemmung, die für das Gelingen des erfindungsgemäßen Verfahrens notwendig sind.

35 Beispiel der Spaltung von Disulfidverbindung im modifizierten NT'. Die Spaltung erfolgt durch eine Zugabe von 20 bis 50mmol/l DTT oder Mercaptoethanol (Sigma) Lösung pH 8 auf die

Reaktionsoberfläche. Die Oberfläche wird 10 min. mit dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

5

Beispiel 2:

Modifiziertes dUTP ($\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$) mit einem kurzen
10 spaltbaren Linker (Fig. 6e-1). Als Ausgangssubstanzen dienen:
Bis-dUTP, synthetisiert nach Hanna (Method in Enzymology 1989,
v.180, S.383), 2-Mercaptoethylamin, MEA, (Sigma).

Zu 400 μl 100mmol/l Bis-dUTP in 40mmol/l Boratpuffer pH 8.5
15 werden 100 μl 100mmol/l MEA-Lösung, pH 8.5, in H_2O zugegeben und
1 Stunde bei RT inkubiert. Anschließend wird zu diesem Gemisch
solange eine gesättigte Lösung an I_2 in 0.3mol/l KI-Lösung zuge-
tropft, bis die Jodfarbe bestehen bleibt. Die Nukleotide (Bis-
dUTP und $\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$) können z.B. durch eine Ethanol-
20 Präzipitation oder auf einer DEAE-Cellulose-Säule in
Ammoniumcarbonat-Gradient (pH 8.5) von anderen
Reaktionsprodukten abgetrennt. Bis-dUTP stört bei der
anschließenden Ankopplung eines Farbstoffs an die Aminogruppe
des Linkers nicht, so dass die Abtrennung des $\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$
25 von bis-dUTP im Endreinigungsschritt erfolgen kann.

In einer ähnlichen Weise kann auch dCTP (Fig.6-e2) modifiziert
werden, dabei dient Bis-dCTP als Ausgangssubstanz
(synthetisiert nach Hanna et al. Nucleic Acid Research 1993,
30 v.21, S.2073).

Weitere NT' (dUTP' und dCTP') mit einem kurzen Linkerrest können
in einer ähnlichen Weise synthetisiert werden, wobei NT'
beispielsweise folgende Strukturen aufweisen können (Fig.6e):

35 $\text{dUTP-SS-(CH}_2)_n\text{-NH}_2$, Fig.6e-1,

$\text{dCTP-SS-(CH}_2)_n\text{-NH}_2$, Fig.6e-2,

wo n zwischen 2 und 6 liegt, vorzugsweise zwischen 2 und 4,

dUTP-SS-(CH₂)_n-X-CO-(CH₂)_m-Z, Fig.6e-3,
dUTP-SS-(CH₂)_n-X-CO-Y-(CH₂)_m-Z, Fig.6e-4,
dCTP-SS-(CH₂)_n-X-CO-(CH₂)_m-Z, Fig.6e-5,
dCTP-SS-(CH₂)_n-X-CO-Y-(CH₂)_m-Z, Fig.6e-6,

5 X = NH, O, S

Y = NH, O, S

Z = NH₂, OH, Farbstoff

wo (n + m) zwischen 4 und 10 liegt, vorzugsweise zwischen 4 und 6.

10

An den Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Waggoner Method in Enzymology 1995 v.246, S.362, Jameson et al. Method
15 in Enzymology 1997, v.278, S.363).

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung des FluoroLink™ Cy3 monofunktional dye (Amersham Pharmacia biotech) (NT*-Struktur Fig. 6i) angegeben.
20 Das ist ein monofunktionaler NHS-Ester-Fluoreszenzfarbstoff. Die Reaktion wird nach Angaben des Herstellers durchgeführt: Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 300 µl 100mmol/l Natrium-Borat-Puffer pH 8.5 aufgelöst. Dazu wird Farbstoff (300nmol) gegeben und 1h bei RT inkubiert. Die
25 Reinigung des mit dem Farbstoff modifizierten NT* erfolgt über RP-HPLC in einem Methanol-Wasser Gradienten.

Ein weiteres Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung von TRITC (Tetramethylrhodamin-5-
30 isothiocyanat, Molecular Probes) angegeben (dUTP-SS-TRITC Fig.6h).

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 30 µl 100mmol/l Natrium-Borat-Puffer pH 9 aufgelöst
35 (10mmol/l NT*). Dazu werden 10 µl 10mmol/l TRITC in DMF gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT* erfolgt über RP-HPLC in einem Methanol-Wasser

Gradienten.

Das so hergestellte NT* erfüllt die Anforderungen des Einbaus in den DNA-Strang, des Fluoreszenznachweises und Kettenabbruchs nach dem Einbau und der Aufhebung der Hemmung, 5 die für das Gelingen des erfindungsgemäßen Verfahrens notwendig sind.

Beispiel der Spaltung der Disulfidverbindung im modifizierten NT'. Die Spaltung erfolgt durch Zugabe von 20 bis 50mmol/l 10 Dithiothreitol-Lösung (DTT) oder Mercaptoethanol-Lösung (Sigma), pH 8, auf die Reaktionsoberfläche. Die Oberfläche wird 10 min. mit dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

15

Weitere Beispiele für NT-Strukturen, die im erfindungsgemäßen Verfahren verwendet werden können sind in Fig. 6k,l,m abgebildet. Für die einzelnen Syntheseschritte siehe beispielsweise J.L.Ruth et al. Molecular Pharmacology 1981 v.20 20 S.415, L. Ötvös et al. NAR 1987 v.15 S.1763, G.E.Wright et al. Pharmac Ther. 1990 v.47, S.447, „Nucleotide Analogs; Synthesis and Biological Function“ K.H. Scheit 1980, Wiley-Interscience Publication, "Nucleic acid chemistry" Ed. L.B.Townsend, v.1-4, Wiley-Interscience Publication, "Chemistry of Nucleosides and 25 Nucleotides" Ed. L.B.Townsend, v.1-3, Plenum Press.

Beispiel 3:

30 Sequenzanalyse mit 4 markierten NTs'

Bei einer bevorzugten Ausführungsform der Erfindung werden alle vier in die Reaktion eingesetzten NTs' mit Fluoreszenzfarbstoffen markiert.

35

Dabei verwendet man eine der oben genannten farbigen Kodierungsschemata. Die Zahl der ermittelten NTs für jede

Sequenz aus einem Genprodukt liegt zwischen 5 und 100, idealerweise zwischen 20 und 50. Diese ermittelten Sequenzen werden mit Hilfe eines Programms mit bekannten Sequenzen in Gen-Datenbanken verglichen und entsprechenden Genen zugeordnet.
5 Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Ein Zyklus hat folgende Schritte:

- 10 a) Zugabe einer Lösung mit markierten Nukleotiden (NTs') und Polymerase zu immobilisierten Nukleinsäureketten,
- b) Inkubation der immobilisierten Nukleinsäureketten mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- 15 c) Waschen
- d) Detektion der Signale von einzelnen Molekülen,
- e) Entfernung der Markierung von den eingebauten Nukleotiden,
- f) Waschen.

20

Beispiel 4:

Sequenzanalyse mit 2 markierten NTs' und 2 unmarkierten NTs (2NTs' / 2NTs-Methode).

25

In einer anderen Ausführungsform werden für die Analyse der Sequenzen 2 modifizierte NTs' und 2 unmodifizierte NTs eingesetzt.

30 Diese Ausführungsform beruht auf dem Prinzip, dass eine Abfolge aus 2 Signalen (markierte NT's) genügend Informationen zur Identifizierung einer Sequenz enthalten kann. Die ermittelte Sequenz wird mit der Referenzsequenz verglichen und einer bestimmten Position zugeordnet, z.B.:

35

ACCAAAACACCC - ermittelte Sequenz (dCTP' und dATP' sind markiert)

ATCATCGTTTCGAAATATCGATCGCCTGATGCC - Referenzsequenz

A-C---C-AAA-A-C-A-C-CC (zugeordnete ermittelte Sequenz)
5 ATCATCGTTTCGAAATATCGATCGCCTGATGCC (Referenzsequenz)

Vorzugsweise werden die ermittelten Sequenzen mit Hilfe eines Programms der Referenzsequenz zugeordnet. Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde
10 liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Die Genprodukte werden wie oben beschrieben zur Sequenzierung vorbereitet und mit 2NTs'/2NTs-Methode sequenziert. Man erhält
15 Sequenzenabschnitte aus Genprodukten, wobei jede Sequenz eine Abfolge aus 2NTs' darstellt. Bekannte Gensequenzen dienen als Referenzsequenzen. Um eine eindeutige Zuordnung der ermittelten Sequenz zu einer bekannten Referenzsequenz zu ermöglichen, muß diese Abfolge lang genug sein. Vorzugsweise beträgt die Länge
20 der ermittelten Sequenzen mehr als 20 NT's. Da 2 markierte NTs' nur einen Teil der Sequenz darstellen, ist die Gesamtlänge des synthetisierten komplementären Strangs ca. doppelt so lang, wie die Abfolge der detektierten NTs' (bei 20 detektierten NTs' beträgt die Gesamtlänge z.B. durchschnittlich 40 NTs).

25

Zur Synthese eines komplementären Stranges werden 4 Nukleotide benötigt. Da die mit einem Fluoreszenzfarbstoff markierten NTs' in der vorliegenden Erfindung als Semiterminatoren auftreten, d.h. die Termination ausschließlich bei Verfügbarkeit modifizierter NTs' auftritt, müssen unmodifizierte NTs in einem zusätzlichen Schritt in jedem Zyklus in die Reaktion zugegeben werden. Die genaue Position dieses Schrittes in dem Zyklus kann variieren. Wichtig dabei ist, dass die markierten NTs' und die unmodifizierte NTs getrennt verwendet werden.

35

Ein Zyklus bei dieser Ausführungsform kann beispielhaft folgendermaßen aussehen:

- a) Zugabe einer Lösung mit modifizierten NTs' und Polymerasen auf die Oberfläche mit den bereitgestellten Genprodukten
- b) Inkubation der immobilisierten Nukleinsäureketten mit
5 dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- c) Waschen
- d) Detektion der Signale von einzelnen, modifizierten und in die den Genprodukten komplementären neusynthetisierten
10 Strängen eingebauten NTs'-Molekülen
- e) Entfernung der Markierung und der terminierenden Gruppe bei den eingebauten Nukleotiden
- f) Waschen
- g) Zugabe von 2 unmodifizierten NTs und Polymerasen
- 15 h) Waschen.

Beispiel 5:

Die Vorbereitung und die Durchführung der Sequenzierungsreaktion.

20

Die Vorbereitung der Gel-Oberfläche:

Das Polyacrylamid-Gel für die Analyse von Reaktionen mit einzelnen Molekülen wird nach allgemeinen Regeln der Gel-Vorbereitung für elektrophoretische Auftrennung erstellt
25 („Electrophoresis“ A.T. Andrews, Oxford science publications 1995).

Die Polymerisationsreaktion kann z.B. durch UV-Licht oder durch Radikalbildner durchgeführt werden. In diesem Beispiel wird Ammoniumpersulfat (APS) und TEMED
30 (Tetramethylethyldiamin) zur Radikalreaktion verwendet, z.B. TEMED 0.01% v/v und APS 0.04% w/v. Die Komponentenzusammensetzung kann breit variieren, die Konzentrationen einzelner Komponenten liegen in folgenden Bereichen (errechnet für die gebrauchsfertige wässrige AA-
35 bisAA-Lösung):

Acrylamid-Monomer (AA) von 3 bis 30%, idealerweise zwischen 10 und 20%

Bis-Acrylamid (bis-AA) im Verhältnis zum Acrylamid-Monomer 1:10 bis 1:50, vorzugsweise 1:20.

Zur Herstellung werden 2 saubere Glasplatten verwendet (mit 5 Aceton und danach Wasser gewaschen). Eine Glasplatte (P1) wird vorzugsweise mit einem wasserabweisenden Reagenz vorbehandelt, z.B. Repel-silan, Dimethyldichlorsilane-Lösung, Amersham Pharmacia-Biotech. P2 dient als fester Träger für das Gel und kann mit gelbindenden Reagenzien z.B. Bind-silan, 10 Methacryloxypropyltrimethoxysilane, Amersham Pharmacia-Biotech, vorbehandelt werden, so dass es zu einer kovalenten Bindung zwischen dem Gel und der Glasoberfläche kommt. Die P2-Vorbehandlung mit gelbindenden Reagentien ist dann sinnvoll, wenn mehrere Reaktionen mit immobilisierten Molekülen 15 durchgeführt werden müssen. Bei einer geringeren Anzahl an Reaktionen ist eine solche Vorbehandlung nicht notwendig. In diesen Fällen reicht, für P2 eine saubere Glas-Oberfläche aus, so dass das Gel allein durch adhäsive Kräfte an der Glasoberfläche haften bleibt.

20

Die fertige Polymerisationslösung (AA/bisAA-Lösung mit Radikalbildnern) wird zwischen P1 und P2 gegossen, so dass eine Schicht mit der Dicke von ca. 5 bis 30 μm resultiert. Die Dicke des Gels kann z.B. durch Abstandhalter kontrolliert 25 werden. Nach Erhärtung wird P1 entfernt. Das Gel bleibt auf P2 haften. Es wird mit entionisiertem Wasser gewaschen.

Das Gel kann direkt weiter verwendet werden oder in verschiedenen Fertigungsstadien getrocknet und gelagert 30 werden. Vor einer Reaktion mit markierten Molekülen wird das Gel normalerweise einige Minuten in der Reaktions-Pufferlösung aufgequollen und erst dann für die Reaktion eingesetzt.

Auf eine so vorbereitete Gel-Oberfläche werden 35 Nukleinsäureketten durch Austrocknen immobilisiert.

Beispielsweise wurde eine Lösung (ca. 1 μl) einer Plasmid-DNA (mit Hind III linearisierte, durch Hitze in einzelsträngige

Form überführte pMOS-Blue-Plasmid-DNA ca. 3400 NT lang, Konzentration $0.1\mu\text{g}/\mu\text{l}$) auf ca. 10mm^2 der Gel-Oberfläche aufgetropft und bei 90°C zum Trocknen gebracht. Die errechnete Dichte der immobilisierten Plasmid-Moleküle betrug ca. 1000 pro $1\mu\text{m}^2$.

Als Primer wurde das Oligonukleotid 5'-AGTGAATTCGAGCTCGGTAC-3' verwendet. Die Primerbindungsstelle (nachfolgend fettgedruckt) zusammen mit der für die Analyse relevanten Verlängerung hat folgende Sequenz:

10 5'-ATCCCCGGGTACCGAGCTCGAATTCACT-3'

Eine Flow-Cell (Mikroflüssigkeitskanal, MFK) mit der Reaktionsoberfläche als Deckel wurde zusammengebaut. Ein solcher MFK erlaubt einen schnellen Flüssigkeitsaustausch unter der Geloberfläche.

Als Vorversuch wurde der Primer (errechnete T_m $45,3^\circ\text{C}$, $0.1\mu\text{mol/l}$ in 50mmol/l Tris-HCl pH 8,7) bei 45°C für 10 Minuten mit der Plasmid-DNA auf der Oberfläche hybridisiert (Annealing). Nach einem Waschschrift wurde die Dichte der Plasmid-Primer-Komplexe kontrolliert. Die Kontrolle erfolgte durch den Einbau von dCTP-Cy3 (Amersham Pharmacia Biotech) unter Verwendung von Klenow-Fragment (2Units pro $50\mu\text{l}$ in 20mmol/l Tris-HCl-Puffer, pH 8,5, mit 5mmol/l MgCl_2 , 15 Minuten bei 30°C). Dabei wird nur ein einzelnes dCMP-Cy3 in den wachsenden Strang eingebaut.

Als Detektionsapparatur diente Axioplan 2e (Zeiss) mit der CCD-Kamera AxioCam (Zeiss), Fig.7,8.

30 Die Signaldichte der einzelnen, eingebauten dCMP-Cy3-Moleküle entspricht der Dichte der extensionsfähigen Plasmid-Primer-Komplexe. Unter den genannten Bedingungen betrug die Dichte der Plasmid-Primer-Komplexe durchschnittlich ca. 15 pro $100\mu\text{m}^2$ und lag damit in der gewünschten Größenordnung (Fig. 8a-c).

35 Auf einer zweiten, in gleicher Weise vorbereiteten Oberfläche (mit Hind III linearisierte, durch Hitze in einzelsträngige Form überführte pMOS-Blue-Plasmid-DNA ca. 3400 NT lang,

Konzentration $0.1\mu\text{g}/\mu\text{l}$ mit hybridisierten Primern) wird eine zyklische Sequenzierungsreaktion durchgeführt. Dabei werden dUTP-SS-CH₂CH₂NH-R-Cy3 (dUTP*) und dCTP-SS-CH₂CH₂NH-R-Cy3 (dCTP*) (s. Beispiel 2) als reversible Terminatoren verwendet.

5 Die Detektionsapparatur ist dieselbe wie im Vorversuch.

Die für die zyklische Sequenzierungsreaktion verwendeten Lösungen setzen sich wie folgt zusammen:

a) Reaktionslösung für die Einbaureaktion: 20 mmol/l Tris-HCl-
10 Puffer, pH 8.5, 5mmol/l MgCl₂, 10% Glycerin, Klenow-Fragment (Amersham Pharmacia-Biotech) 2U pro 50 μl , dUTP* bzw. dCTP*, oder dATP und dGTP je 10 $\mu\text{mol/l}$.

b) Waschlösung: 20 mmol/l Tris-HCl pH 8.5, 0.01% Na-Azid

c) Reaktionslösung für die Abspaltungsreaktion: 20 mmol/l
15 Tris-HCl, pH 8.5, 50mmol/l Mercaptoethanol.

Die Einbaureaktionen mit markierten NT*s wurden bei 30° C 15 Minuten durchgeführt.

20 Im ersten Zyklus der Sequenzierungsreaktion wurde eine Reaktionslösung mit dCTP* zugegeben. Nach einem Waschschrift wurde ein Detektionsschritt durchgeführt, wobei Einzelmolekül-Signale mit den zugeordneten x,y-Koordinaten auf der Oberfläche registriert wurden (insgesamt ca. 11.200 Signale).
25 Danach wurde die Markierung von den eingebauten NT*s abgespalten (Raumtemperatur, 10 Minuten) und die Oberfläche gewaschen.

Im zweiten Zyklus wurde eine Reaktionslösung mit dUTP*
30 zugegeben und 15 Minuten lang bei 30° C inkubiert. Nach einem anschließenden Waschschrift wurden die Einzelmolekül-Signalen auf der Oberfläche detektiert (insgesamt ca. 200 Signale). Dies entspricht dem Hintergrundsignal, das durch eine unspezifische Bindung der NT*s an die Oberfläche entsteht. Die
35 Markierung von den NT*s wurde abgespalten (Raumtemperatur, 10 Minuten) und die Oberfläche mit der Waschlösung gewaschen.

Im dritten Zyklus wurde eine Reaktionslösung mit dATP und dGTP

zugegeben und 15 Minuten lang bei 30° C inkubiert. Anschließend wurde die Oberfläche gewaschen.

Die Zyklen 1 bis 3 wurden drei mal wiederholt, wobei insgesamt 5 ca. 9900 CCU-Zielsequenzen ermittelt wurden. Diese Sequenzen können eindeutig dem Primer zugeordnet werden.

Legenden zu Figuren 1 bis 9**Legende zu Fig. 1****5 Schematische Darstellung der Analyse von mRNA-Population**

Der Analyse liegt die Sequenzierung kurzer Abschnitte von mRNA zugrunde.

1) mRNA - die zu analysierende mRNA-Population, in diesem
10 Beispiel bestehend aus zwei unterschiedlichen mRNA-Molekülpopulationen (dünne und dicke Streifen repräsentieren mRNA-Moleküle)

2) Immobilisierte mRNA - an die plane Oberfläche gebundene
15 mRNA-Primer-Komplexe, in diesem Beispiel erfolgt die Bindung durch die Oligo-dT-Primer

3) Zugabe einer Lösung mit Polymerasen und NT*s - der erste
20 Schritt in einem Zyklus der Sequenzierungsreaktion

4) Waschschrift - nach dem Einbauschrift wird die Oberfläche gewaschen

5) Detektion - die Signale von einzelnen eingebauten NT*s
25 werden detektiert

6) Entfernung der Markierung und der zur Termination führenden Gruppe - zur Fortsetzung der Sequenzierungsreaktion werden die Markierung und das sterische Hindernis entfernt
30

Legende zur Fig. 2

Beispiel für allgemeine Struktur von Genprodukt-Primer-
35 Komplexen

Fig. 2 In dieser Ausführungsform hat jedes Genprodukt eine

einheitliche Primerbindungsstelle (PBS) an seinem 3'-Ende, so dass an diese PBS ein einheitlicher Primer binden kann.

- 1) Primer
- 2) Primerbindungsstelle
- 3) Einzelsträngiger Teil des Genprodukts, der analysiert wird

Legende zur Fig. 3

10 Ein Beispiel für die Erzeugung einer einheitlichen Primerbindungsstelle (PBS).

3a) In diesem Fall werden NTs an das 3'-Ende der einzelsträngigen Genprodukte angekoppelt (ein so genanntes
15 „Tailing“). Durch Verwendung eines einheitlichen NT entsteht eine einheitliche PBS.

3b) Nach Hybridisierung von Primern entstehen Genprodukt-Primer-Komplexe

20

Legende zur Fig. 4

Beispiel für die Bindung von Genprodukten an eine gelartige
25 Reaktionsoberfläche.

Auf einer festen Unterlage (1) haftet eine Gelschicht (2), z.B. ein Polyacrylamidgel (Fig. 4a), oder haften viele Gelkügelchen (5), z.B. Agarosekügelchen (Fig. 4b). An die
30 Oberfläche des Gels sind Genprodukte (4) gebunden. Die Genprodukte tragen eine funktionelle Gruppe, z.B. Biotin, und sind an das Gel über Streptavidin oder Avidin (3) gebunden.

35 Legende zur Fig. 5

Beispiel für eine Durchflussvorrichtung

Eine gelartige Reaktionsoberfläche (1) ist auf einer für das Anregungs- und Fluoreszenzlicht durchlässigen festen Unterlage (2) befestigt. Sie bilden zusammen den Deckel der Flow-Cell.
5 Die Flüssigkeiten in der Flow-Cell können kontrolliert ausgetauscht werden, wobei die Flow-Cell zusammen mit Vorratsbehälter (3), Pumpe (4) und Ventil (5) eine Durchflussvorrichtung bilden. Auf der Reaktionsoberfläche sind Genprodukt-Primer-Komplexe gebunden (hier nicht abgebildet).
10 Die Signale der eingebauten NT*s werden mit der Detektionsapparatur (6) detektiert.

Legende zur Fig. 6

15 Strukturen von 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können.

Fig. 6a - Schematische Darstellung der NT-Struktur, bei der die spaltbare Gruppe und die sterisch anspruchsvolle, zur
20 Termination führende Gruppe Teile des Linkers bilden. Der Linker ist die Verbindung zwischen Nukleobase und Fluoreszenzfarbstoff.

A,B,C,D,E - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination
25 führende Gruppe, F - Fluoreszenzfarbstoff.

Fig. 6b - Schematische Darstellung der NT-Struktur, wobei die spaltbare Gruppe ein Teil des Linkers ist und der Fluoreszenzfarbstoff gleichzeitig die sterisch anspruchsvolle,
30 zur Termination führende Gruppe darstellt.

A,B,C - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination führende Gruppe, F - Fluoreszenzfarbstoff.

35 Fig. 6c - Schematische Darstellung der Struktur von eingebauten NT*s nach dem Abspaltungsschritt. Dargestellt sind zwei NT*s mit dem verbliebenen Linkerrest (A).

Fig. 6d - Schematische Darstellung der NT-Struktur, wobei die spaltbare Gruppe, die gleichzeitig die sterisch anspruchsvolle, zur Termination führende Gruppe ist, einen Teil des Linkers darstellt.

A,B,C,D - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination führende Gruppe, F - Fluoreszenzfarbstoff.

10

Fig. 6e - Darstellung von bevorzugten NT-Strukturen, bei denen der Linker an die 5-Position im Pyrimidinring angekoppelt ist.

Fig. 6f - Darstellung anderer bevorzugter NT-Strukturen, bei denen der Linker an die 5-Position im Pyrimidinring angekoppelt ist.

Fig. 6g - Darstellung von bevorzugten NT-Strukturen, bei denen der Linker an die 7-Position im Purinring angekoppelt ist.

20

Fig. 6h,i,j - Beispiele für die Ankopplung von Farbstoffen an den Linker

Fig. 6k

25 Strukturen von weiteren 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können. Der Linker ist an die 5-Position des Pyrimidinrings gekoppelt.

Die Substituenten $R_{1,2,3,4}$ sind wählbar und können unabhängig 30 voneinander auftreten.

Die Z-Gruppe stellt in einer Ausführungsform (6k-1) die Verbindung zwischen dem Linker und der Base dar. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, 35 Ether-, Thioether- oder Aminogruppe sein.

Die E-Gruppe stellt in einer Ausführungsform (6k-1) einen

internen Teil des Linkers dar. In einer anderen Ausführungsform (6k-2) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine unverzweigte Alkyl- oder Alkenylkette mit einer Zahl von 5 Kohlenstoffatomen, vorzugsweise zwischen 1 und 5, sein.

Die E-Gruppe kann aber auch eine Alkyl- oder Alkenylkette mit einer internen Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung sein.

10 Die C-Gruppe ist in diesem Beispiel eine chemisch spaltbare Gruppe. In den Ausführungsformen (6k-1,2) stellt sie einen internen Teil des Linkers dar. In einer anderen Ausführungsform (6k-3) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann 15 eine Ester-, Thioester- und Disulfidverbindung sein.

Die Y-Gruppe stellt einen internen Teil des Linkers dar, der die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar 20 und kann eine verzweigte oder unverzweigte Alkyl- oder Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine Alkyl- oder Alkenylkette mit einer internen Amid- Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

25

Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F) abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, 30 Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

Fig. 6L

Strukturen von weiteren 2'-Deoxynukleosidtriphosphaten, die im 35 Verfahren eingesetzt werden können. Der Linker ist an die 7-Position des Purinrings gekoppelt.

Die Substituenten $R_{1,2,3,4}$ sind wählbar und können unabhängig voneinander auftreten.

Die Z-Gruppe stellt in einer Ausführungsform (6L-1) die
5 Verbindung zwischen dem Linker und der Base dar. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

Die E-Gruppe stellt in einer Ausführungsform (6L-1) einen
10 internen Teil des Linkers dar. In einer anderen Ausführungsform (6L-2) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine unverzweigte Alkyl- oder Alkenylkette mit einer Zahl von Kohlenstoffatomen, vorzugsweise zwischen 1 und 5, sein.

15 Die E-Gruppe kann aber auch eine Alkyl- oder Alkenylkette mit einer internen Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung sein.

Die C-Gruppe ist in diesem Beispiel eine chemisch spaltbare
20 Gruppe. In den Ausführungsformen (6L-1,2) stellt sie einen internen Teil des Linkers dar. In einer anderen Ausführungsform (6L-3) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine Ester-, Thioester- und Disulfidverbindung sein.

25

Die Y-Gruppe stellt einen internen Teil des Linkers dar, der die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar und kann eine verzweigte oder unverzweigte Alkyl- oder
30 Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine Alkyl- oder Alkenylkette mit einer internen Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

35 Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F)

abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

5 Fig. 6m

Weitere Beispiele für Strukturen von 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können. Der Linker ist an die 5-Position des Pyrimidinrings gekoppelt.

10

Die Substituenten $R_{1,2,3,4}$ sind wählbar und können unabhängig voneinander auftreten.

Die Y-Gruppe stellt einen internen Teil des Linkers dar, der 15 die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar und kann eine verzweigte oder unverzweigte Alkyl- oder Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine 20 Alkyl- oder Alkenylkette mit einer internen Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung 25 sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F) abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

30

Legende zur Fig. 7

Beispiel für ein Detektionssystem

35 Dargestellt ist ein Weitfeld-Optik-Detektionssystem. Nach dem Einbau von markierten NT*s wird die Oberfläche (7) abgescannt, wobei die Fluoreszenzsignale von einzelnen, an die NTs

gekoppelten Farbstoffmolekülen detektiert werden.

Legende zur Fig. 8

5

Schematische Darstellung des Scannvorganges

Fig. 8a Schematische Darstellung eines Abschnittes der Reaktionsoberfläche (grau), der abgescannt wird. Die Kreise
10 entsprechen jeweils der Aufnahme eines 2D-Bildes und repräsentieren die Flächen, von denen die Fluoreszenzsignale detektiert werden. Dabei werden pro Aufnahme mehrere Signale (beispielsweise 100 bis 10.000) von einzelnen Molekülen gleichzeitig registriert. Die Reaktionsoberfläche wird in
15 jedem Zyklus abgescannt, wobei während des Scannvorganges mehrere Aufnahmen von unterschiedlichen Stellen der Oberfläche gemacht werden. Dabei können bis zu mehreren Millionen Signale von eingebauten NT*s aufgenommen werden. Die hohe Parallelität ist die Grundlage für die Geschwindigkeit des Verfahrens.

20

Fig. 8b Eine Aufnahme (ein 2D-Bild) mit Signalen von einzelnen, eingebauten NT*s. Zur Versuchsbeschreibung siehe Beispiel 5.

25 Fig. 8c Ausschnitt aus Abbildung 8b. Der Ausschnitt zeigt Signale von vier eingebauten NT*s. Jedes Signal besitzt charakteristische Eigenschaften der Einzelmolekülsignale (s. Beschreibung) und kann auf Grund dieser identifiziert werden (vorzugsweise mit Hilfe eines Computer-Programms). Jedem der
30 identifizierten Signale werden die entsprechenden X,Y-Koordinaten zugeordnet.

Legende zur Fig. 9

35

Beispiel einer vorteilhaften Anordnung von Reaktionsoberflächen. Der Durchsatz wird durch Verwendung von

zwei getrennten Flow-Cells (Mikroflüssigkeitskanäle, MFK) erhöht. Während in der einen Flow-Cell biochemische und chemische Reaktionen ablaufen, wird in der anderen die Detektion durchgeführt. Anschließend tauschen die Flow-Cells ihre Positionen.

Patentansprüche:

1. Verfahren zur parallelen Analyse der Genexpression bei dem man

einzelsträngige Genprodukte bereitstellt, man

die Genprodukte unter Verwendung eines einheitlichen oder mehrerer unterschiedlicher Primer in Form von Genprodukt-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der Genprodukte unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den auf der Oberfläche gebundenen Genprodukt-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' strukturell an der Base so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter

Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man

- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NTs' durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) zur Erzeugung unmarkierter, (NTs oder) Genprodukte die Fluoreszenzfarbstoffe und die sterisch anspruchsvollen Liganden von den am komplementären Strang angefügten NTs' abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner Genprodukt-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser Genprodukte durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt und man aus den ermittelten Teilsequenzen die Identität der Genprodukte bestimmt.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass

man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus nur jeweils ein markiertes NT' einsetzt.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils zwei unterschiedlich markierte NTs' einsetzt.
4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils vier unterschiedlich markierte NTs' einsetzt.
5. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass bereits bekannte Gene als Referenzsequenzen dienen und man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs' und zwei unmarkierte NTs einsetzt und man die Identität der Genprodukte durch Vergleich der gewonnenen Sequenzen mit denen der Referenzsequenzen ermittelt.
6. Verfahren nach den Ansprüchen 1 bis 5, dadurch gekennzeichnet, dass man in die Genprodukte jeweils eine Primerbindungsstelle (PBS) einführt, wobei die Primerbindungsstellen für alle Genprodukte jeweils gleiche oder verschiedene Sequenzen aufweisen.
7. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die Genprodukte mit Primern in einer Lösung unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen, und man die gebildeten Genprodukt-Primer-Komplexe anschließend auf der Reaktionsoberfläche bindet.

8. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die Genprodukte zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit Primern unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wobei Genprodukt-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
9. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die Primer zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit Genprodukten unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wodurch Genprodukte an die Oberfläche gebunden und Genprodukt-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
10. Verfahren nach den Ansprüchen 1 bis 9, dadurch gekennzeichnet, dass die Dichte der extensionsfähigen Genprodukt-Primer-Komplexe zwischen etwa 10 und 100 pro $100\mu\text{m}^2$ liegt.
11. Verfahren nach den Ansprüchen 1 bis 10, dadurch gekennzeichnet, dass der zur strukturellen Modifikation verwendete sterisch anspruchsvolle Ligand der zur Markierung verwendete Fluoreszenzfarbstoff ist.
12. Verfahren nach den Ansprüchen 1 bis 11, dadurch gekennzeichnet, dass die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.
13. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass

die Kunststoffe Polycarbonate oder Polystyrole oder Derivate derselben sind.

14. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass die Gele Agarose- oder Polyacrylamidgele oder Derivate derselben sind.
15. Verfahren nach Anspruch 14, dadurch gekennzeichnet, dass die Gele 1 bis 2 % Agarose-Gele oder 5 bis 15 % Polyacrylamid-Gele sind.
16. Verfahren nach den Ansprüchen 1 bis 15, dadurch gekennzeichnet, dass die Polymerase eine Polymerase ohne 3'-5'-Exonukleaseaktivität ist.
17. Verfahren nach Anspruch 16, dadurch gekennzeichnet, dass die Polymerase aus der Gruppe bestehend aus thermolabilen viralen, bakteriellen DNA-Polymerasen, thermolabilen oder thermostabilen viralen Reversen Transcriptasen, thermostabilen DNA-Polymerasen oder deren Modifikationen ausgewählt ist.
18. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase Sequenase Version 2 ist.
19. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase Taq-Polymerase ist.
20. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase ProHa-DNA-Polymerase ist.
21. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase Klenow Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' Exonukleaseaktivität ist.
22. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase AMV-Reverse Transcriptase ohne RNase-

Aktivität ist.

23. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase M-MLV Reverse Transcriptase ist.
24. Verfahren nach Anspruch 17, dadurch gekennzeichnet, dass die Polymerase HIV-Reverse Transcriptase ohne RNase-Aktivität ist.
25. Verfahren nach den Ansprüchen 1 bis 24, dadurch gekennzeichnet, dass die Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene und deren Derivaten ausgewählt sind.
26. Verfahren zur quantitativen Analyse der Genexpression, dadurch gekennzeichnet, dass man das Verfahren nach den Ansprüchen 1 bis 25 durchführt und man anschließend die Anzahl der ermittelten Sequenzen für jedes Genprodukt bestimmt.
27. Verfahren nach Anspruch 26, dadurch gekennzeichnet, dass man zusätzlich zu den Genprodukten ferner ein oder mehrere Kontrollsequenzen in bekannten Konzentrationen verwendet, wobei die Gesamtkonzentration der Kontrollsequenzen zwischen 0,01 und 10% der Gesamtkonzentration der zu analysierenden Probe liegt.
28. Verfahren nach den Ansprüchen 1 bis 27, dadurch gekennzeichnet, daß man eine Sequenzanalyse von mindestens etwa 100.000 bis etwa 10.000.000 Genprodukt-Molekülen durchführt.
29. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass auf seiner Oberfläche die Genprodukte in einer zufälligen Anordnung gebunden sind, wobei die Dichte der gebundenen Genprodukt-Moleküle zwischen 10 und 100 pro 100µm² liegt.

30. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass auf seiner Oberfläche die Genprodukte in einer zufälligen Anordnung gebunden sind, wobei die Dichte der gebundenen Genprodukt-Moleküle zwischen 100 und 1.000.000 pro $100\mu\text{m}^2$ liegt.
31. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass auf seiner Oberfläche die Primer in einer zufälligen Anordnung gebunden sind, wobei die Dichte der gebundenen Primer-Moleküle zwischen 10 und 100 pro $100\mu\text{m}^2$ liegt.
32. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass auf seiner Oberfläche die Primer in einer zufälligen Anordnung gebunden sind, wobei die Dichte der gebundenen Primer-Moleküle zwischen 100 und 10.000 pro $100\mu\text{m}^2$ liegt.
33. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass auf seiner Oberfläche die Primer in einer zufälligen Anordnung gebunden sind, wobei die Dichte der gebundenen Primer-Moleküle zwischen 10.000 und 1.000.000 pro $100\mu\text{m}^2$ liegt.
34. Kit zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 28, dadurch gekennzeichnet, dass es eine Reaktionsoberfläche (einen festen Träger), zur Durchführung des Verfahrens erforderliche Reaktionslösungen, ein oder mehrere Polymerasen, und Nukleotide (NTs) enthält, von denen ein bis vier mit Fluoreszenzfarbstoffen markiert sind, wobei die markierten NTs ferner strukturell so modifiziert sind (NT' bzw. NTs'), dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist,

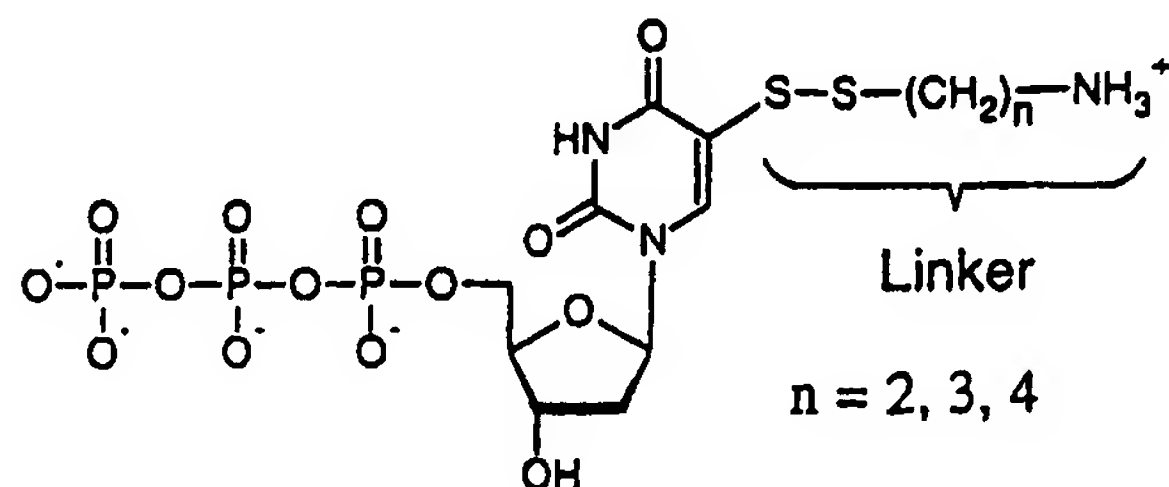
ein weiteres NT^{*} in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist.

35. Kit nach Anspruch 34, dadurch gekennzeichnet, dass der zur strukturellen Modifikation verwendete sterisch anspruchsvolle Ligand der zur Markierung verwendete Fluoreszenzfarbstoff ist.
36. Kit nach Anspruch 34 oder 35, dadurch gekennzeichnet, dass es ferner Bestandteile enthält:
 - a) Oligonukleotid-Primer,
 - b) zur Abspaltung der Fluoreszenzfarbstoffe und sterisch anspruchsvollen Liganden erforderliche Reagenzien, und/oder
 - c) Waschlösungen.
37. Kit nach den Ansprüchen 34 bis 36, dadurch gekennzeichnet, dass die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.
38. Kit nach Anspruch 37, dadurch gekennzeichnet, dass die Kunststoffe Polycarbonate oder Polystyrole oder Derivate derselben sind.
39. Kit nach Anspruch 37, dadurch gekennzeichnet, dass die Gele Agarose- oder Polyacrylamidgele oder Derivate derselben sind.
40. Kit nach Anspruch 39, dadurch gekennzeichnet, dass die Gele 1 bis 2% Agarose-Gele oder 5 bis 15% Polyacrylamidgele sind.
41. Kit nach den Ansprüchen 34 bis 40, dadurch gekennzeichnet,

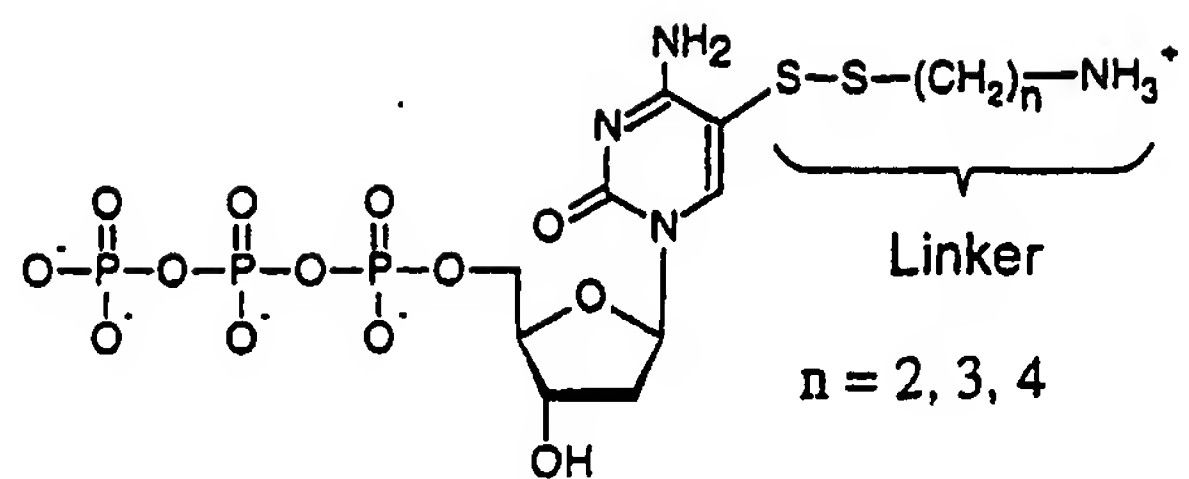
daß die Reaktionsoberfläche ein Träger nach den Ansprüchen 29 bis 33 ist.

42. Kit nach den Ansprüchen 34 bis 41, dadurch gekennzeichnet, dass die DNA-Polymerase eine DNA-Polymerase ohne 3'-5'-Exonukleaseaktivität ist.
43. Kit nach Anspruch 42, dadurch gekennzeichnet, dass die Polymerase aus der Gruppe bestehend aus viralen DNA-Polymerasen vom Sequenase-Typ, bakteriellen DNA-Polymerasen, Reversen Transcriptasen und thermostabilen DNA-Polymerasen ausgewählt ist.
44. Kit nach Anspruch 43, dadurch gekennzeichnet, dass die Polymerase Sequenase Version 2, Klenow Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' Exonukleaseaktivität, AMV-Reverse Transcriptase ohne RNase-Aktivität, M-MLV Reverse Transcriptase, HIV-Reverse Transcriptase ohne RNase-Aktivität, Taq-Polymerase oder ProHa-DNA-Polymerase ist.
45. Kit nach den Ansprüchen 34 bis 44, dadurch gekennzeichnet, dass die Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene und deren Derivaten ausgewählt sind.
46. Nukleotide der Formel

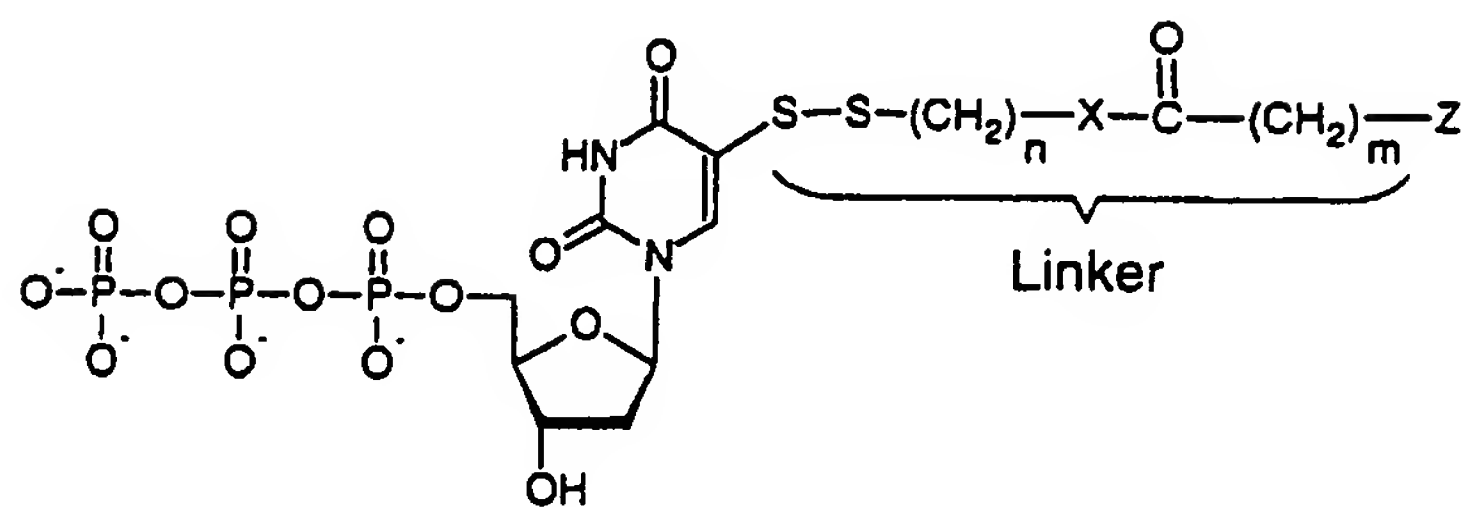
6e-1)



6e-2)

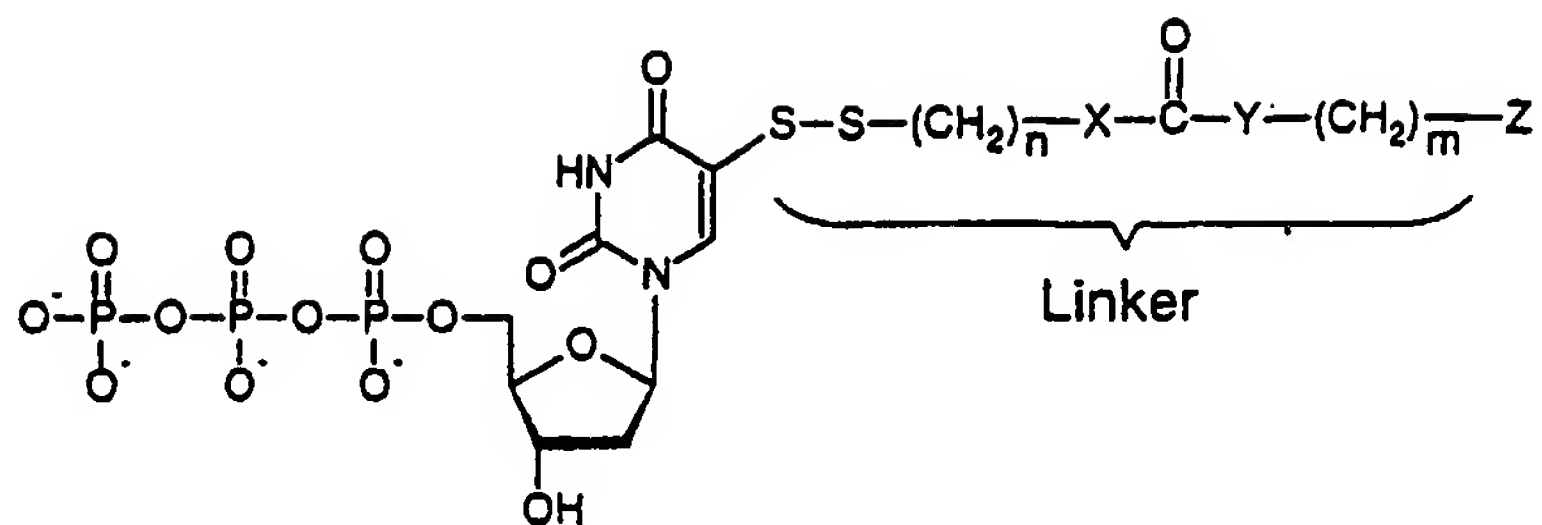


6e-3)

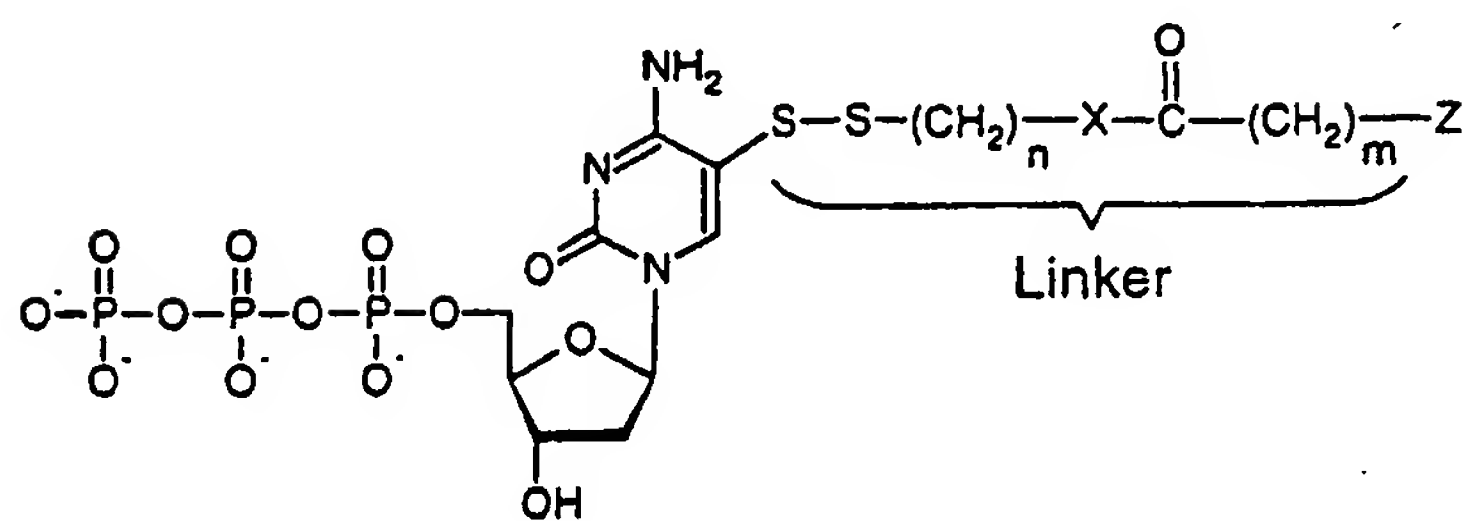


$\text{X} = \text{NH, O, S}$
 $\text{Z} = \text{NH}_2, \text{OH, Farbstoff}$

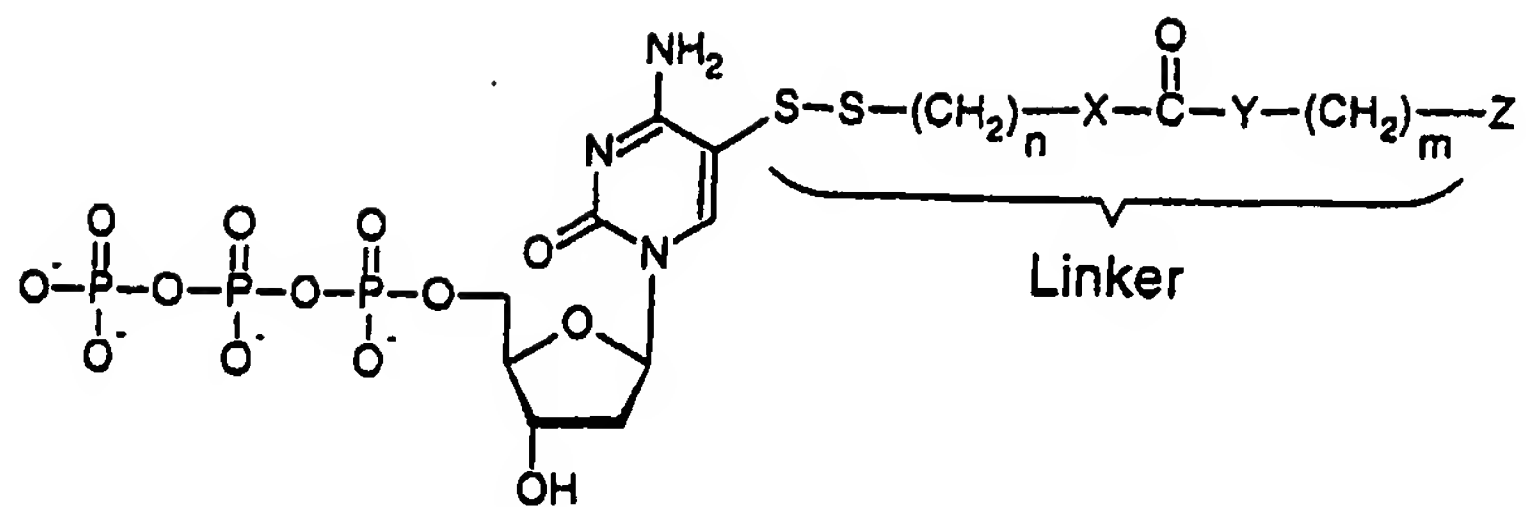
6e-4)

 $X = NH, O, S$ $Y = NH, O, S$ $Z = NH_2, OH$

6e-5)

 $X = NH, O, S$ $Z = NH_2, OH, \text{Farbstoff}$

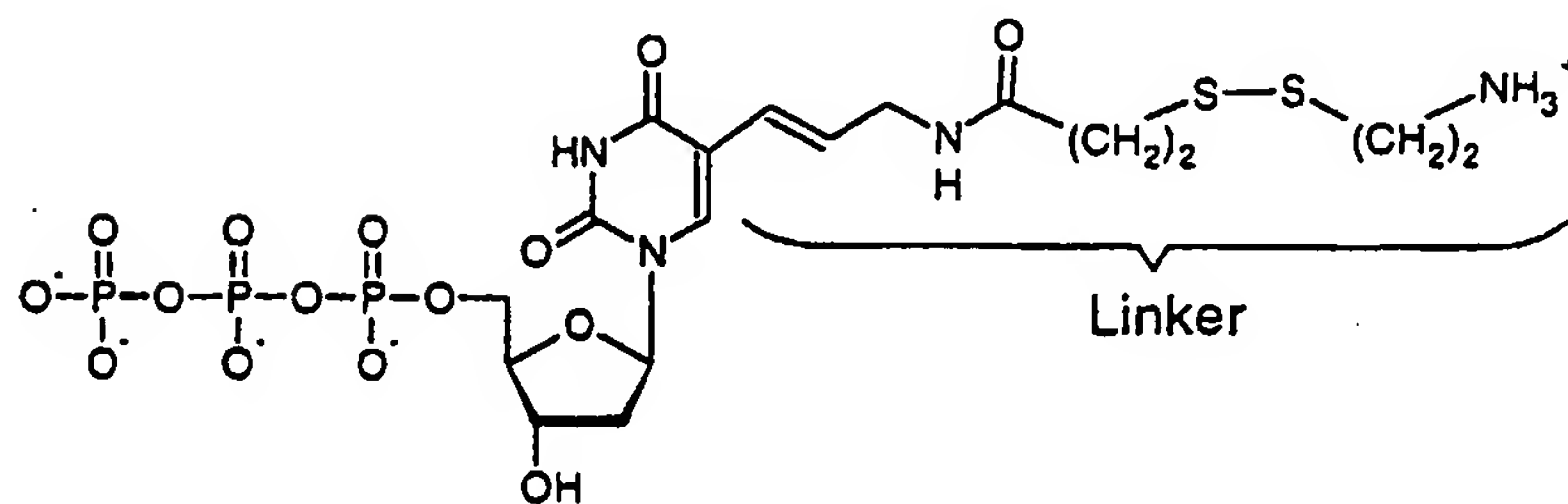
6e-6)



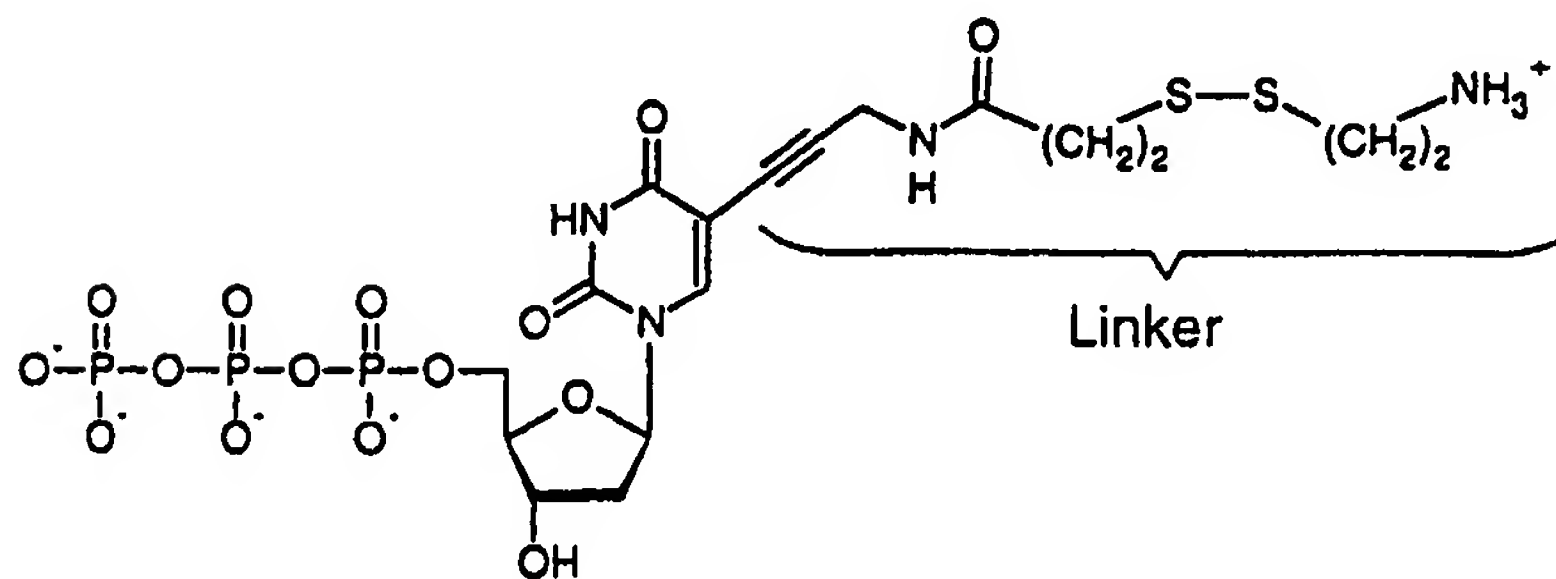
$\text{X} = \text{NH, O, S}$
 $\text{X} = \text{NH, O, S}$
 $\text{Z} = \text{NH}_2, \text{OH}$

47. Nukleotide der Formel

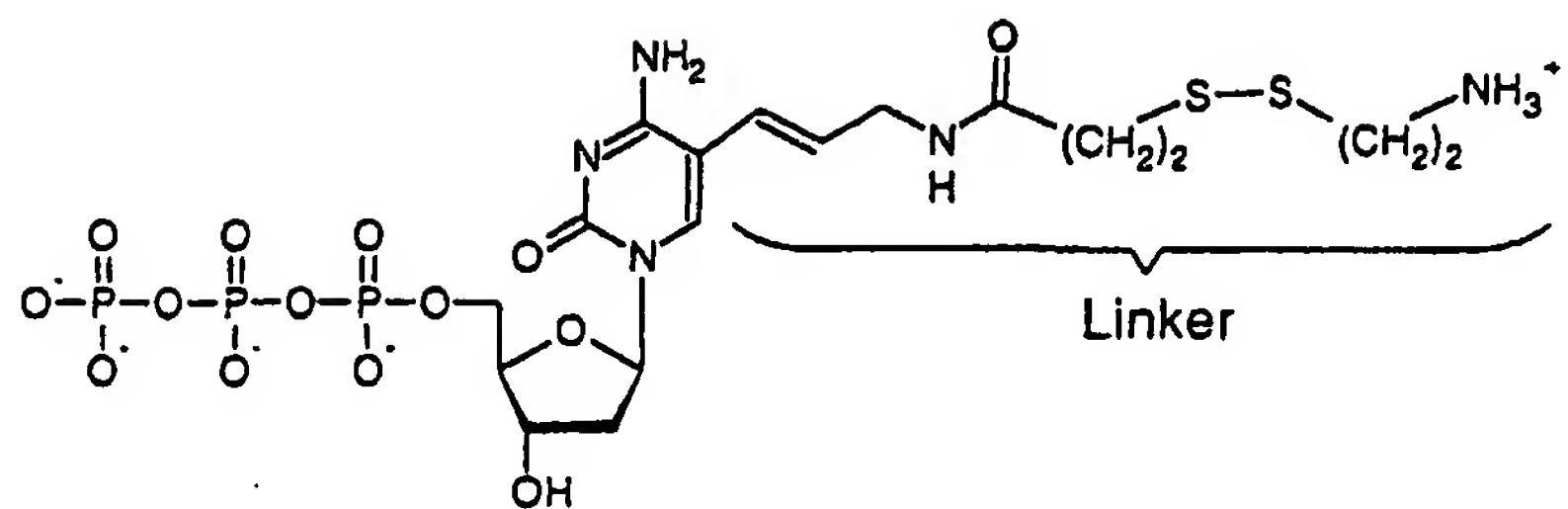
6f-1)



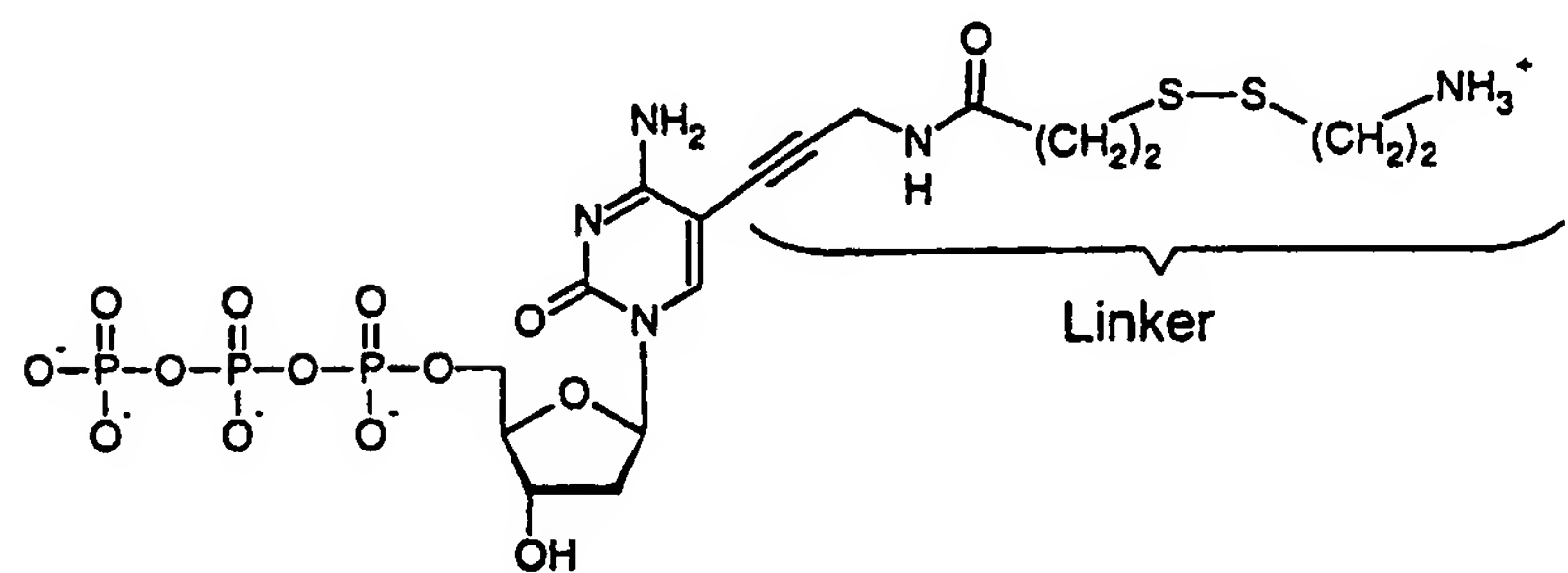
6f-2)



6f-3)

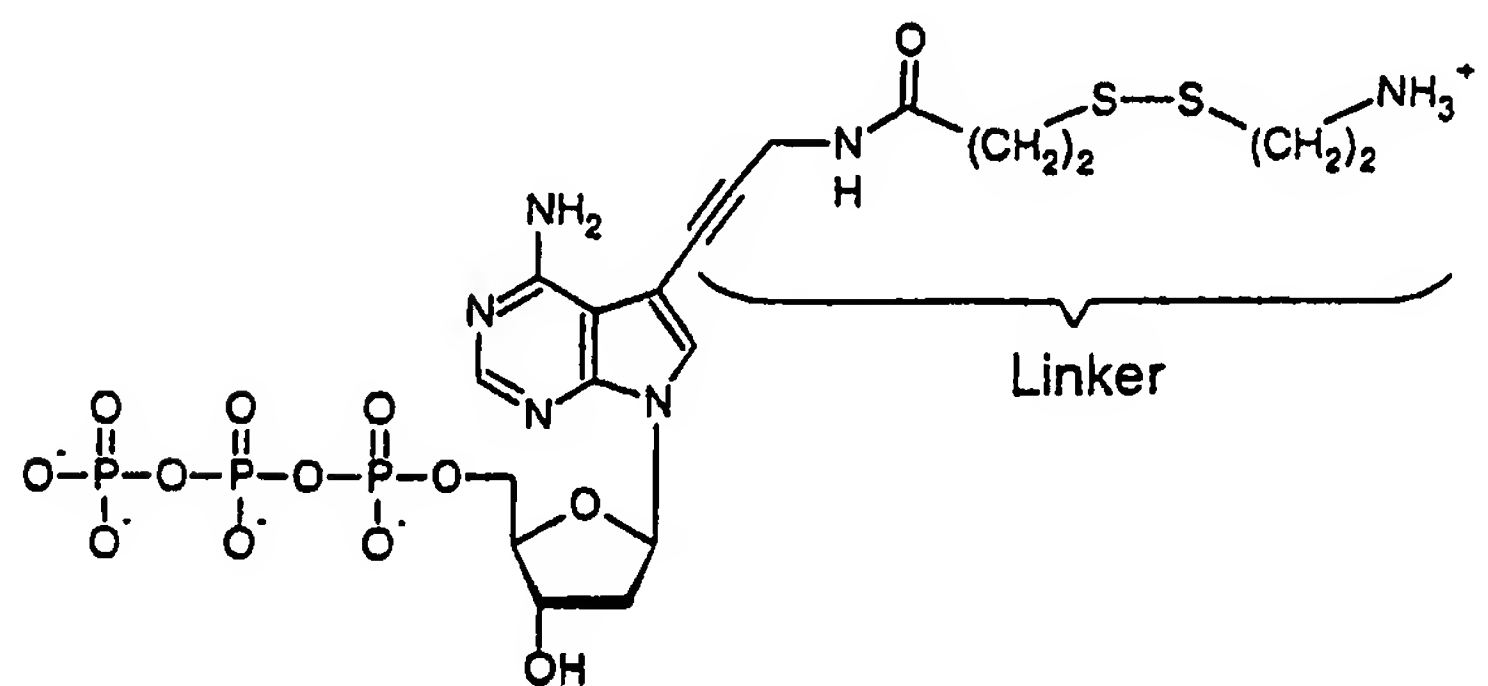


6f-4)

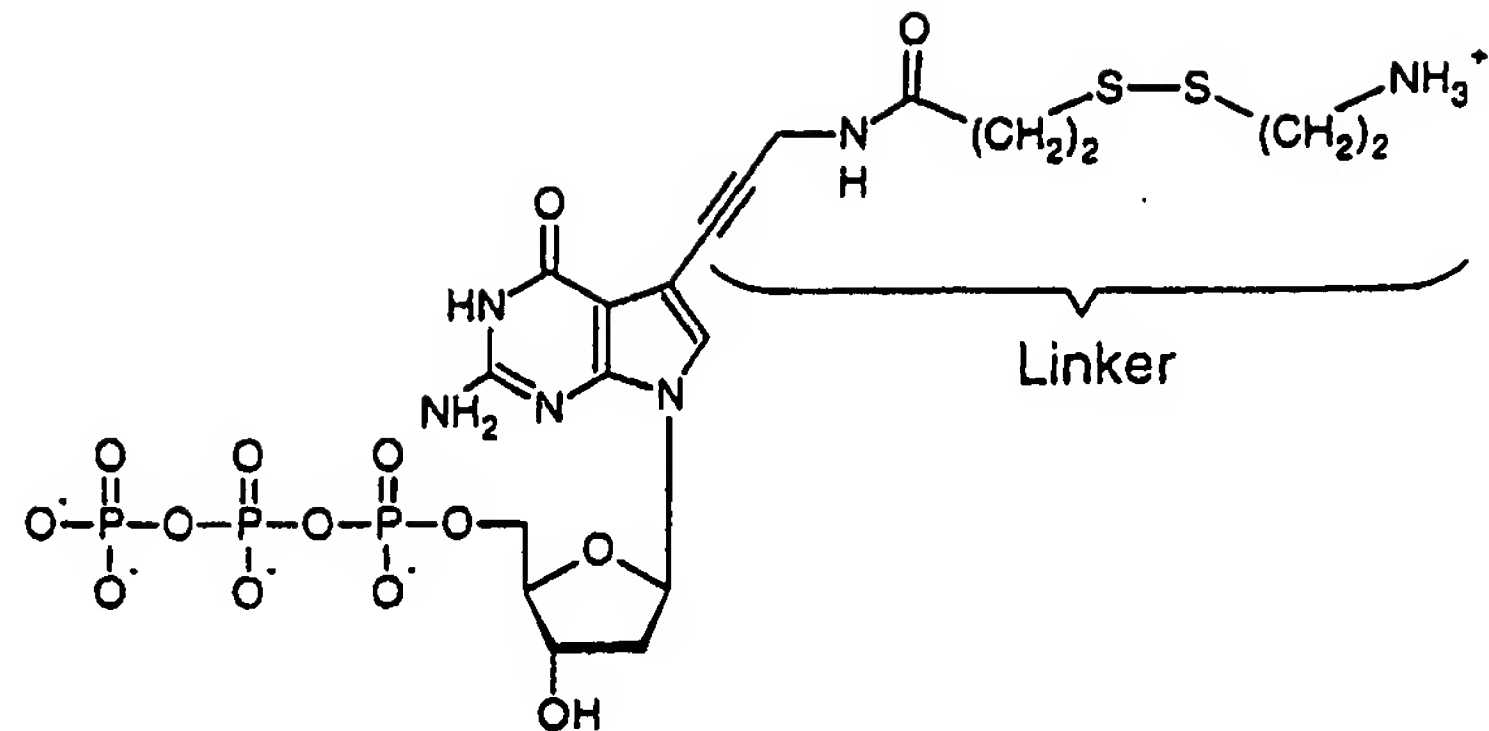


48. Nukleotide der Formel

6g-1)



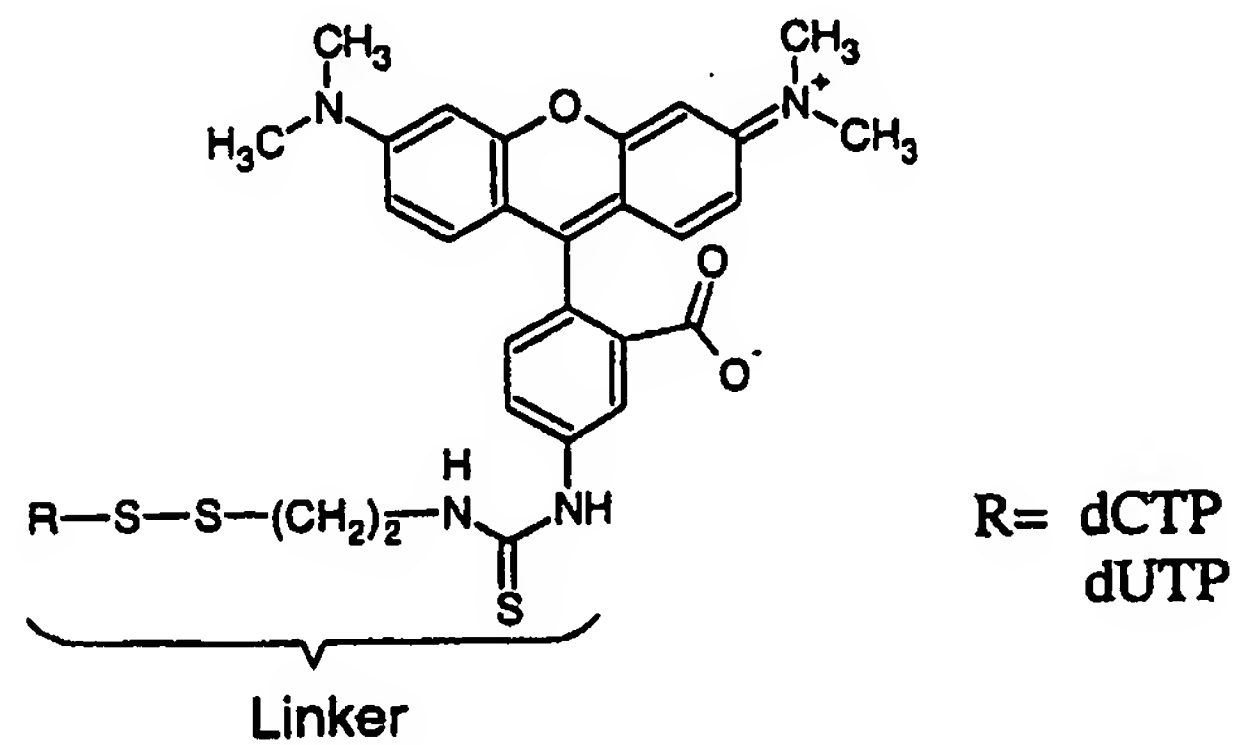
6g-2)



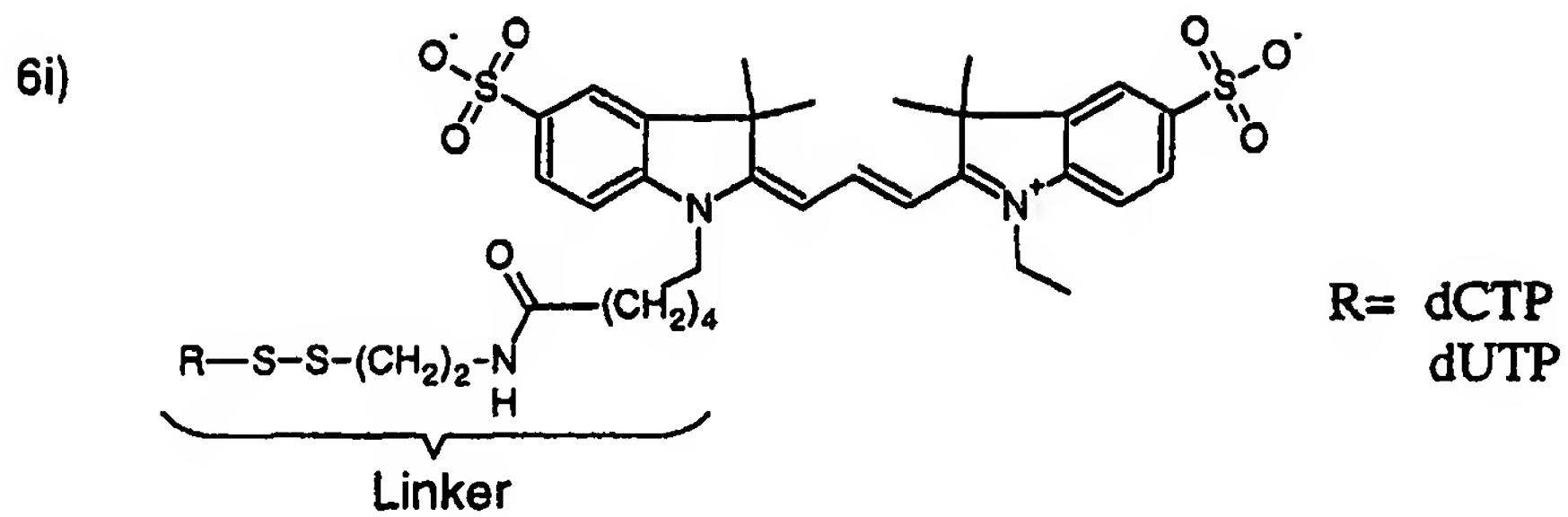
49. Nukleotid nach Anspruch 46 bis 48, dadurch gekennzeichnet, dass an die terminale Aminogruppe ein Fluoreszenzfarbstoff angeheftet ist.

50. Nukleotid der Formel

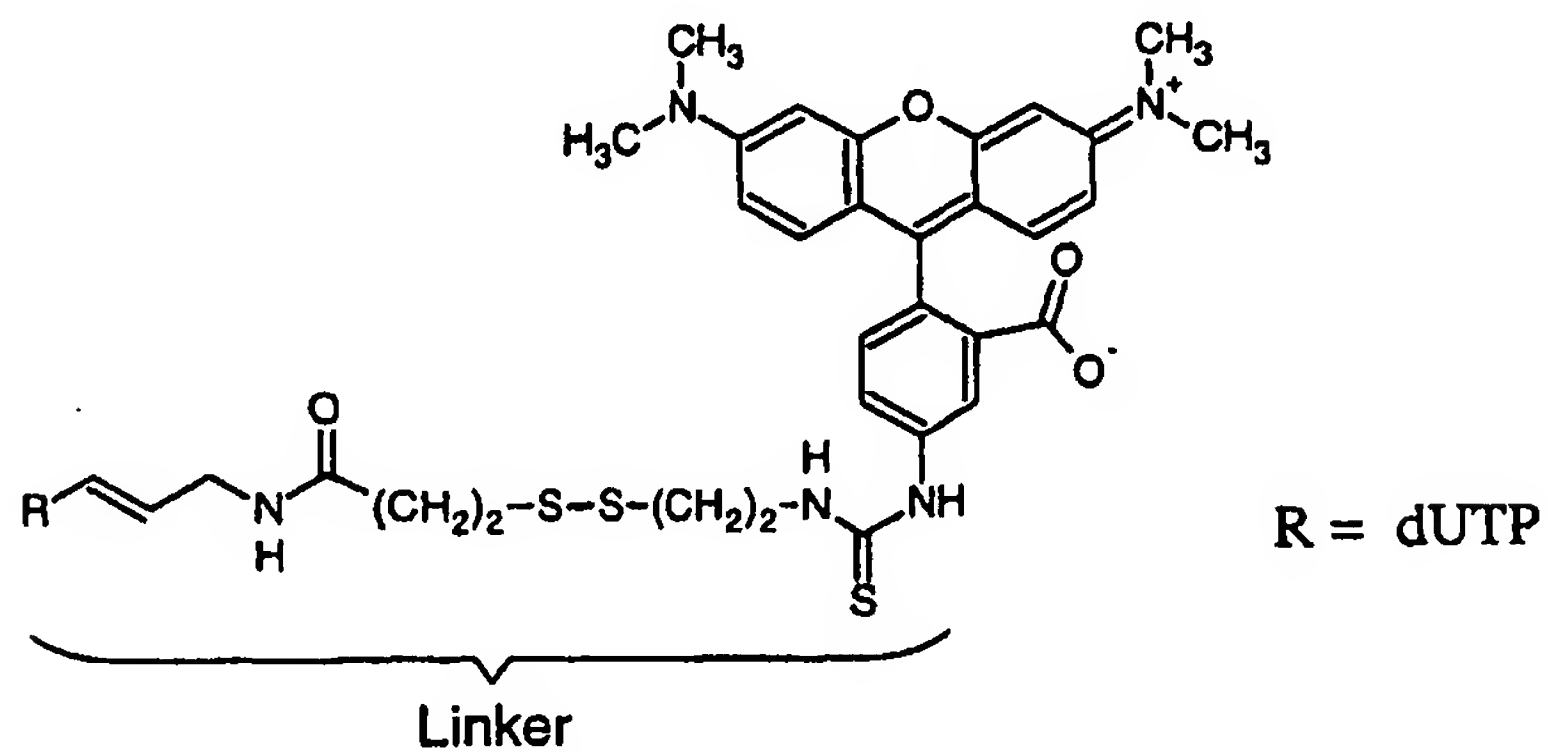
6h)



51. Nukleotid der Formel



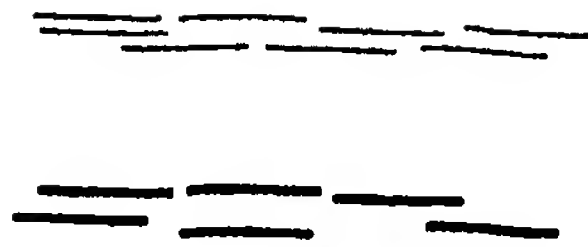
52. Nukleotid der Formel

6j) dNTP-SS-TRITC (L 14)

53. Verwendung eines Nukleotids nach den Ansprüchen 46 bis 52 als NT' in einem Verfahren nach den Ansprüchen 1 bis 28.

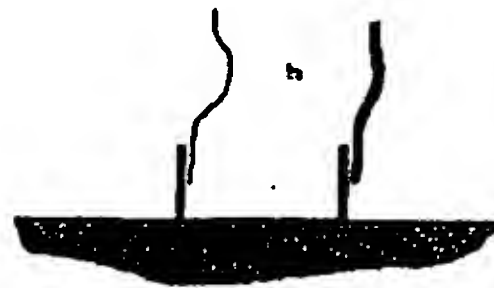
54. Verwendung eines Nukleotids entsprechend Fig. 6k, 6L, 6m, als NT' in einem Verfahren nach den Ansprüchen 1 bis 28.

1)



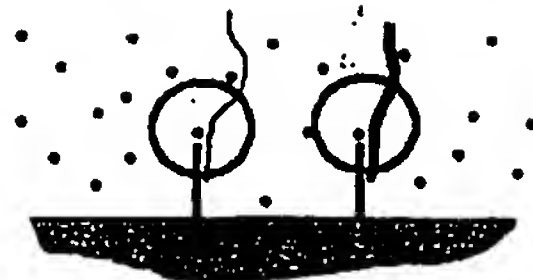
mRNA

2)



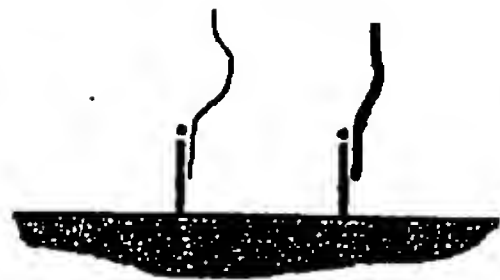
Immobilisierte mRNA

3)



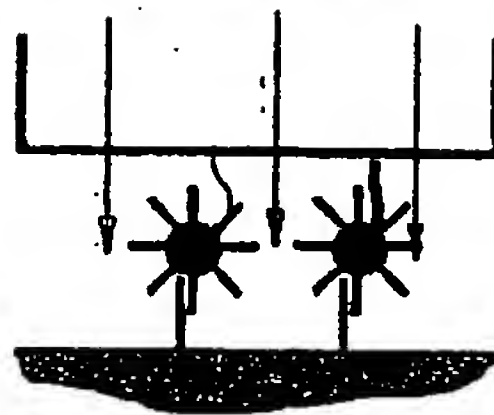
Zugabe einer Lösung mit
Polymerasen und NTs*

4)



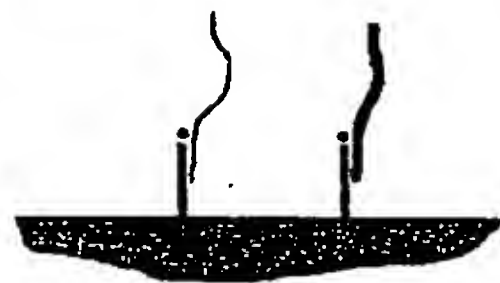
Waschschritt

5)



Detektion

6)



Entfernung der Markierung und
der zur Termination führenden
Gruppe

Legende:

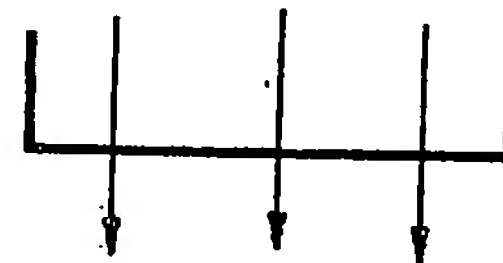
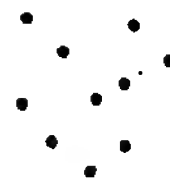
feste Oberfläche
mit oligo-dT-Primern

Polymerase

NT*

Signal von NT*

Mikroskop und
Beleuchtung



Allgemeine Struktur für immobilisierte Genprodukte**Legende:**

- 1) Primer
- 2) Primerbindungsstelle
- 3) einzelsträngiger Teil des Genprodukts, stammt vom Ausgangsmaterial

Nukleotid-Tailing

4a)

5'-----3' einzelsträngiges DNA-Fragment

+ Deoxynucleotidyltransferase
+ dATP

↓

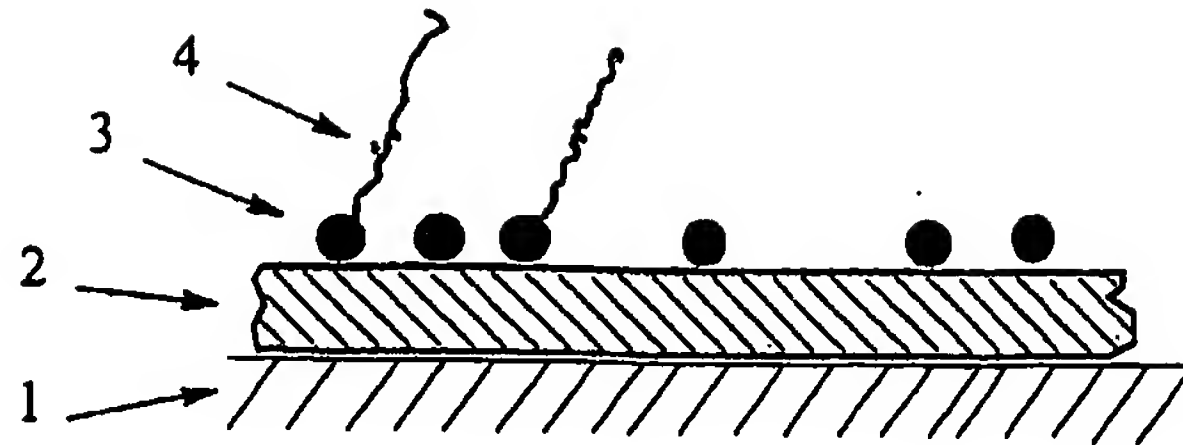
5'------(A)_n3'

4b)

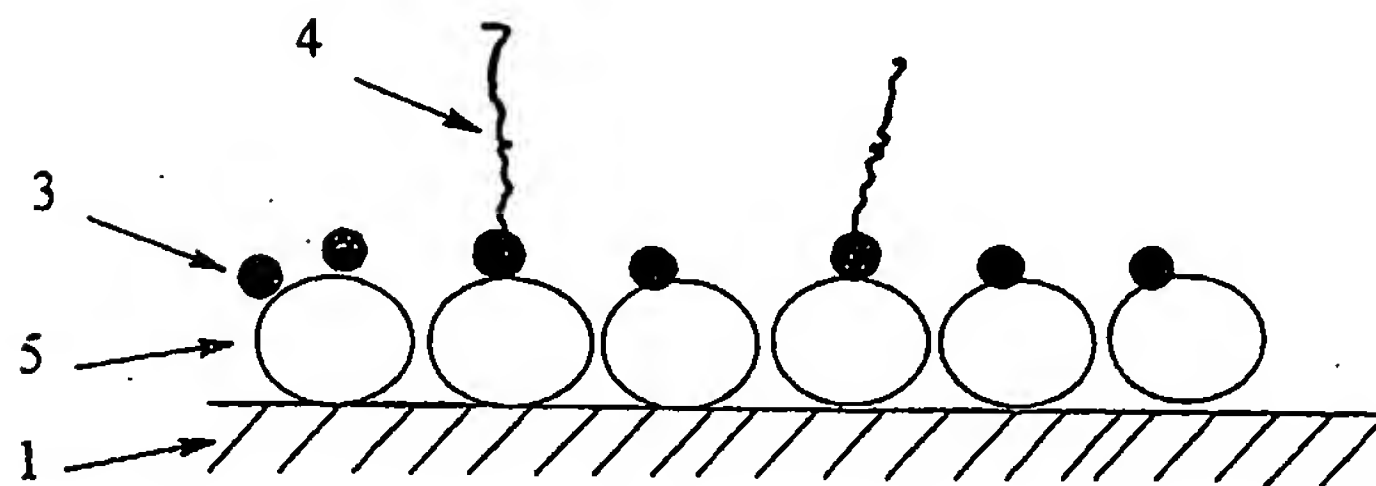
Nach der Hybridisierung mit dem Primer entsteht folgende Konstruktion

5'-----XY(A)_n3'
NM(T)_x5'

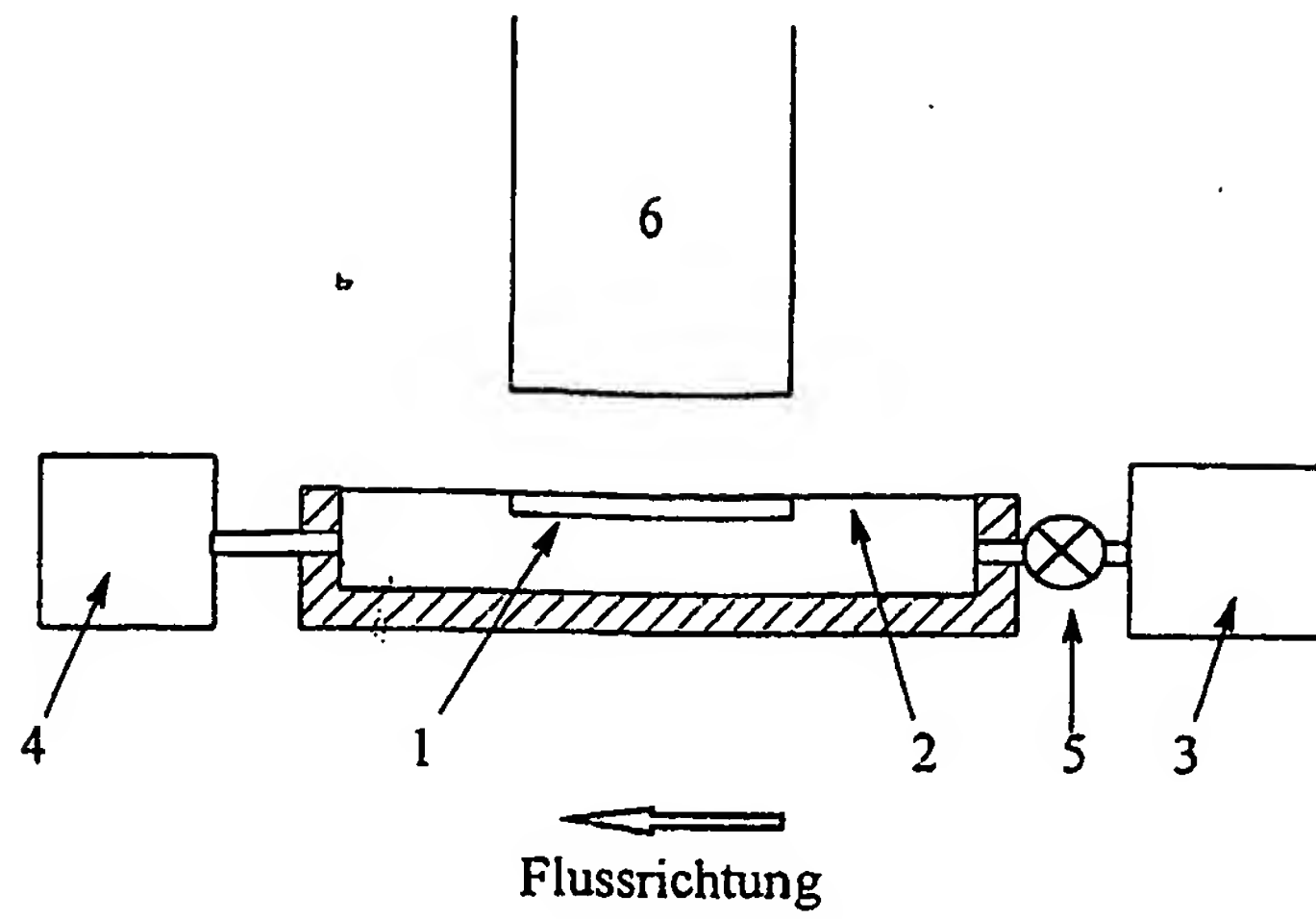
4a)



4b)

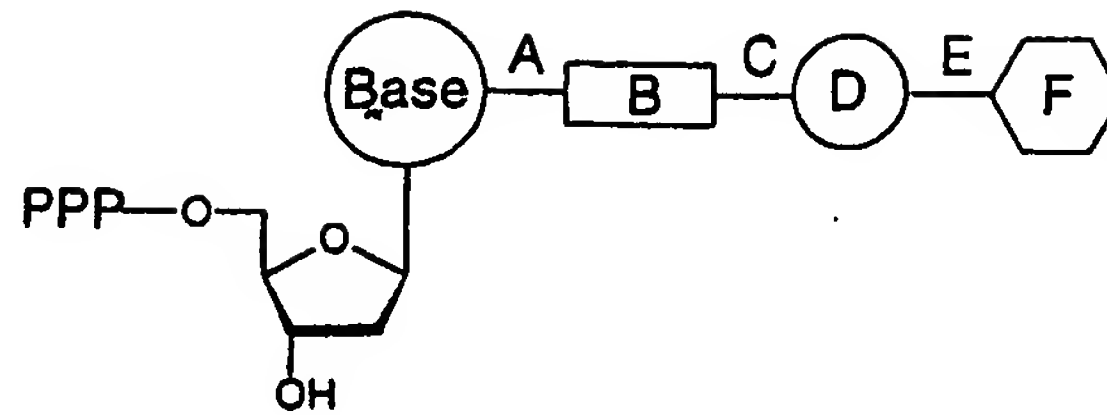
**Legende:**

- 1) feste Oberfläche
- 2) Gel
- 3) zur Immobilisation geeignete Gruppen, z.B. Avidin oder Streptavidin, oder Oligonukleotide
- 4) Genprodukte
- 5) Agarosekügelchen

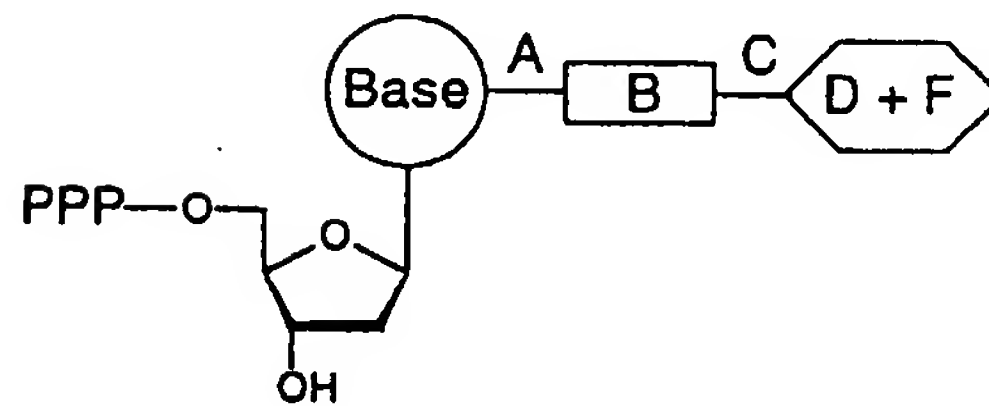
**Legende:**

- 1) Reaktionsoberfläche
- 2) feste Unterlage
- 3) Vorratsbehälter
- 4) Pumpe
- 5) Ventil
- 6) Objektiv des Mikroskops

6a)

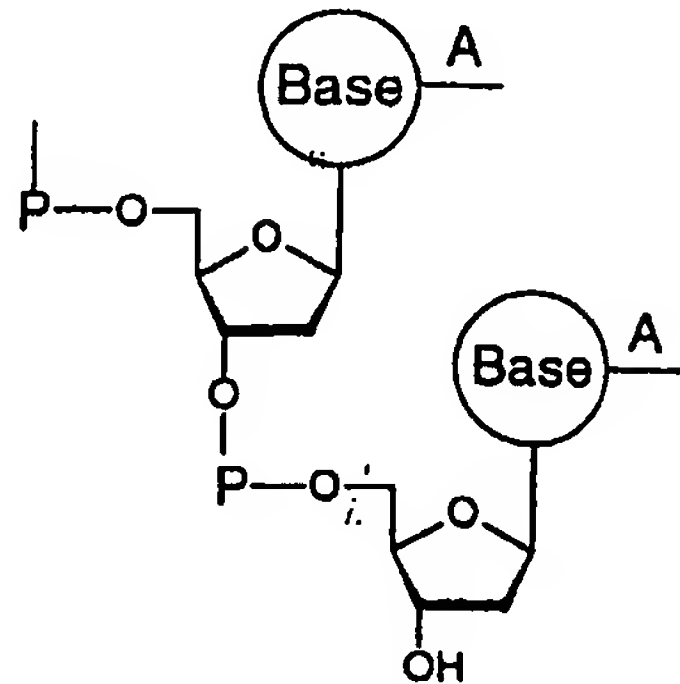


6b)

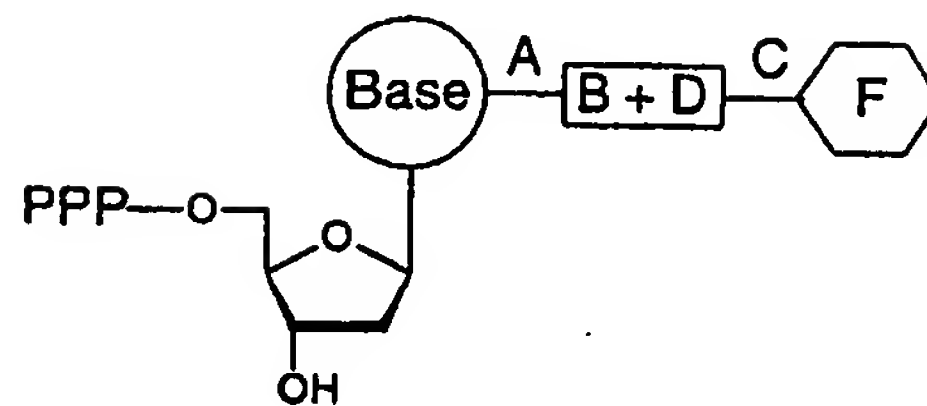
**Legende:**

- A, C, E – Verbindungselemente im Linker
- A – Linkerrest nach der Spaltung
- B – spaltbare Verbindung / Gruppe
- D – zur Termination führende Gruppe
- F – Fluoreszenzfarbstoff (Marker)

6c) eingebaute modifizierte NTs* nach der Spaltung



6d)



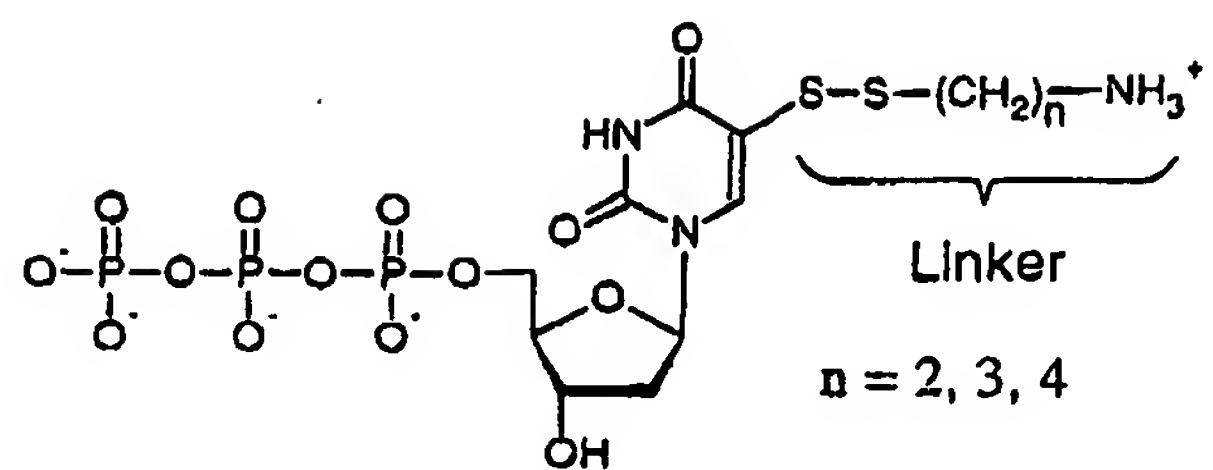
Legende:

- A, C, E – Verbindungselemente im Linker
- A – Linkerrest nach der Spaltung
- B – spaltbare Verbindung / Gruppe
- D – zur Termination führende Gruppe
- F – Fluoreszenzfarbstoff (Marker)

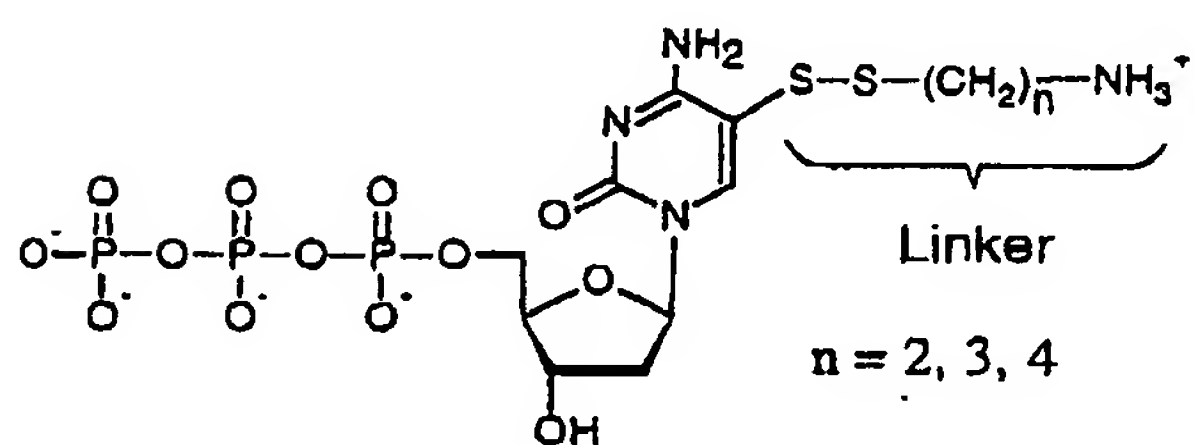
8/20

Fig. 6

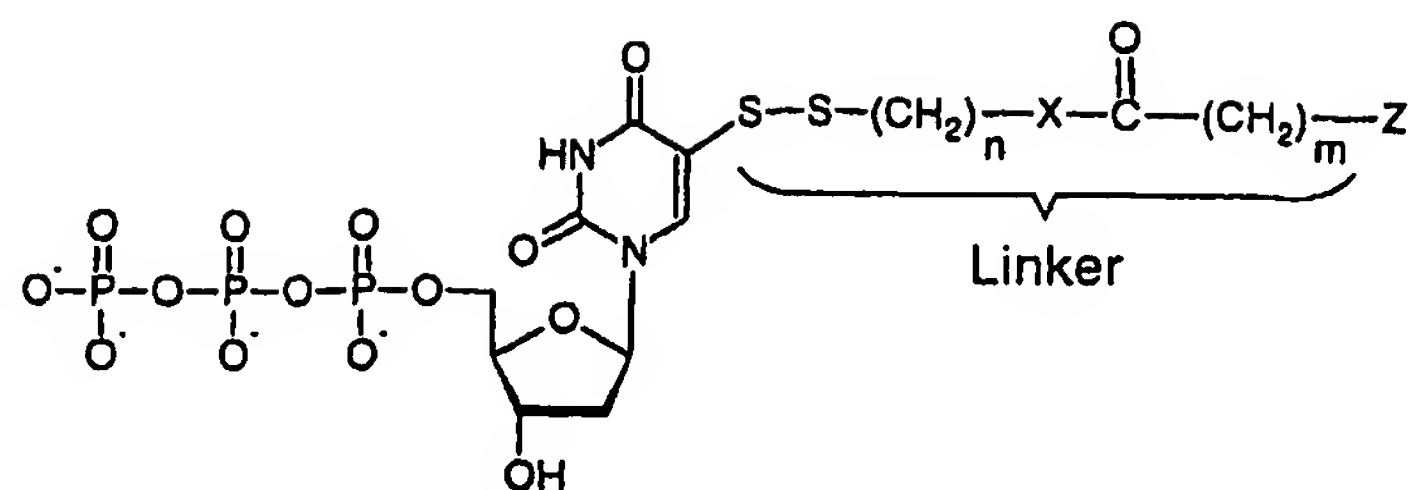
6e-1)



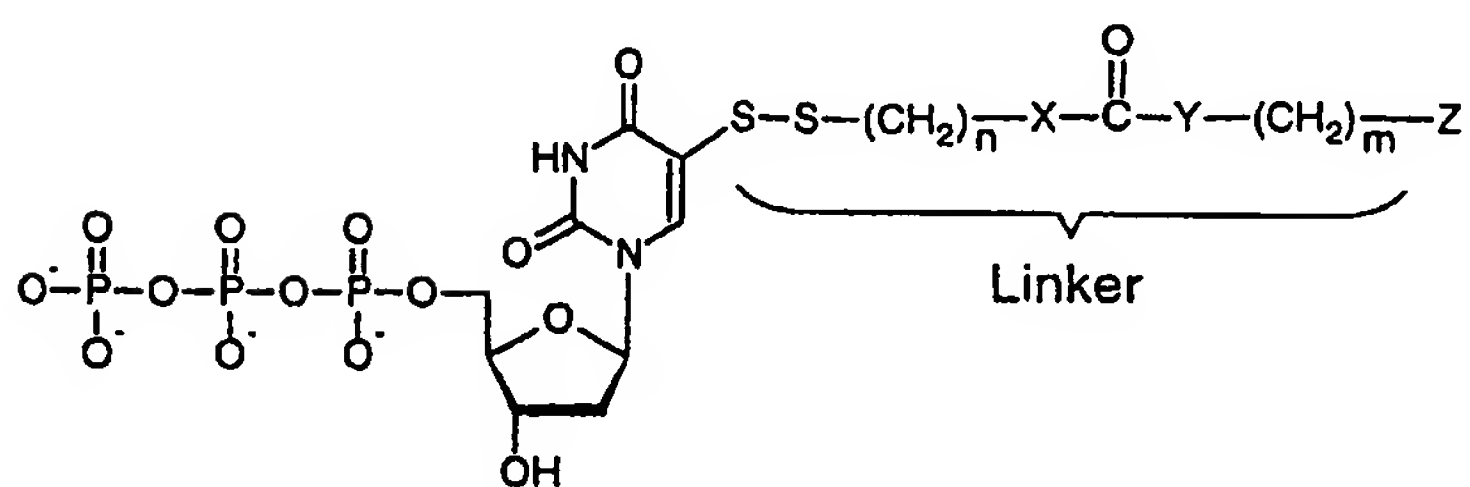
6e-2)



6e-3)

 $X = \text{NH}, \text{O}, \text{S}$ $Z = \text{NH}_2, \text{OH}, \text{Farbstoff}$

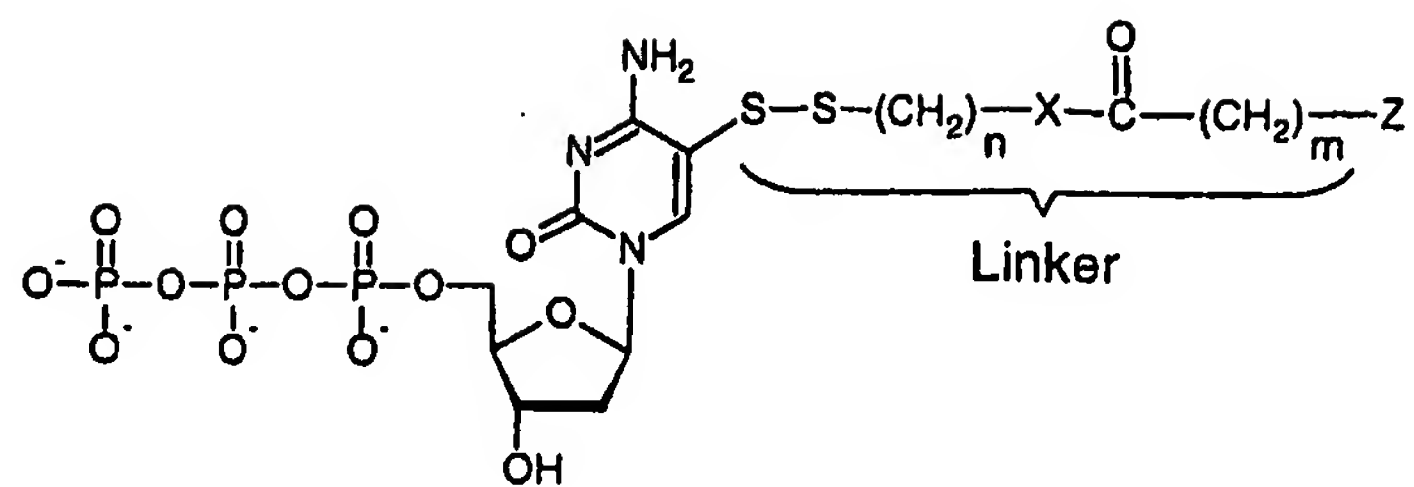
6e-4)

 $X = \text{NH}, \text{O}, \text{S}$ $Y = \text{NH}, \text{O}, \text{S}$ $Z = \text{NH}_2, \text{OH}$

10/20

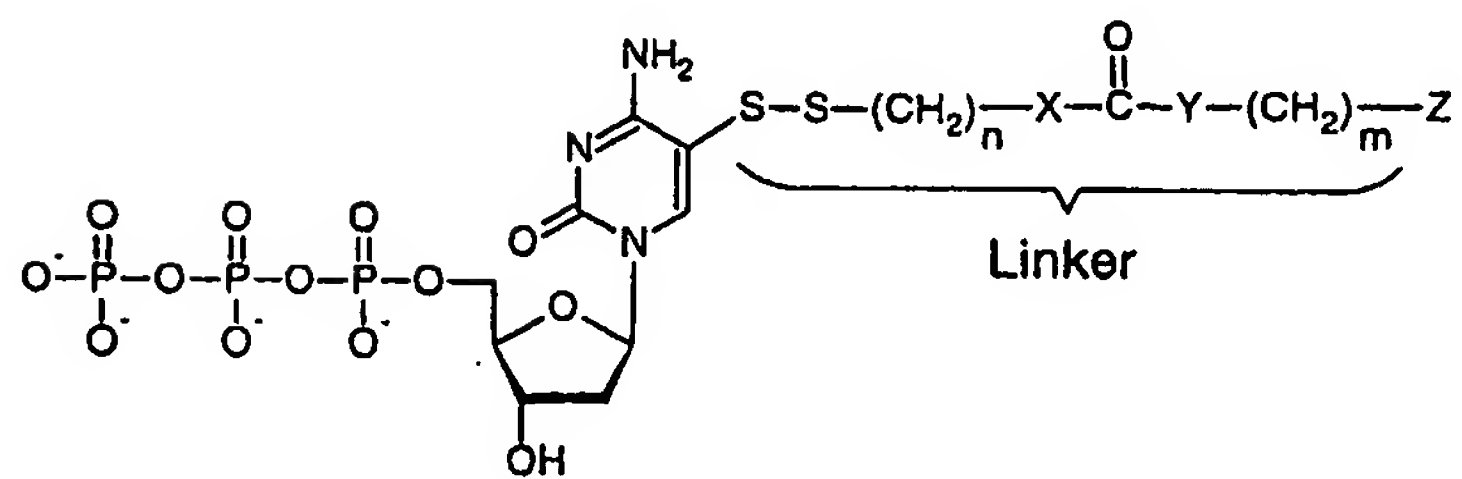
Fig. 6

6e-5)



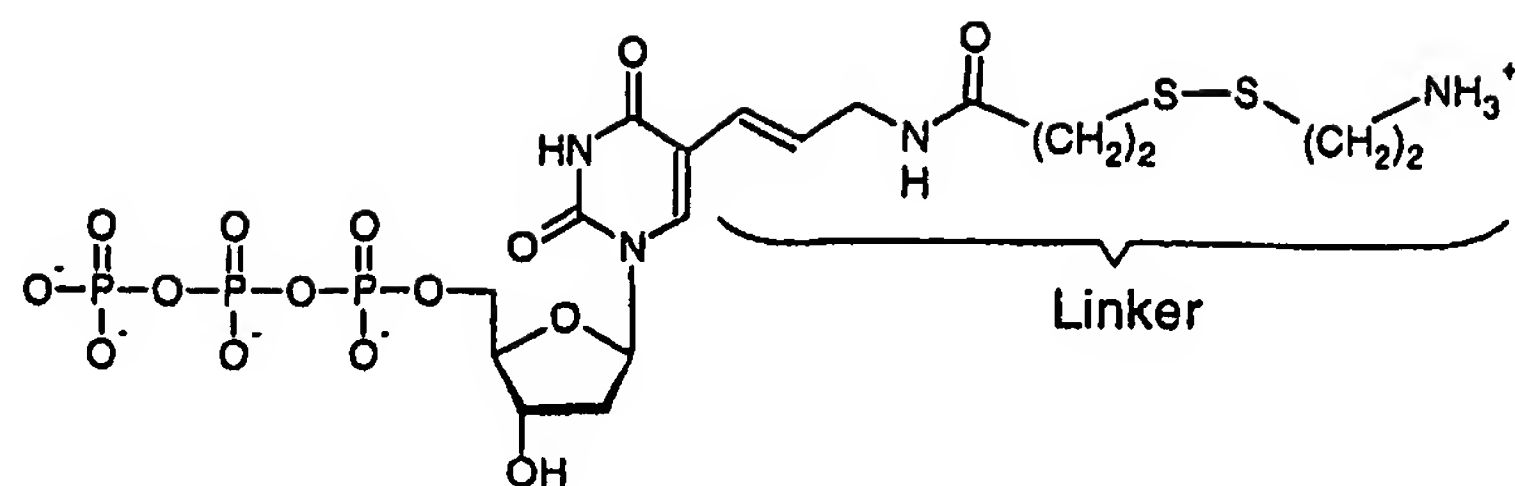
$X = \text{NH}, \text{O}, \text{S}$
 $Z = \text{NH}_2, \text{OH}, \text{Farbstoff}$

6e-6)

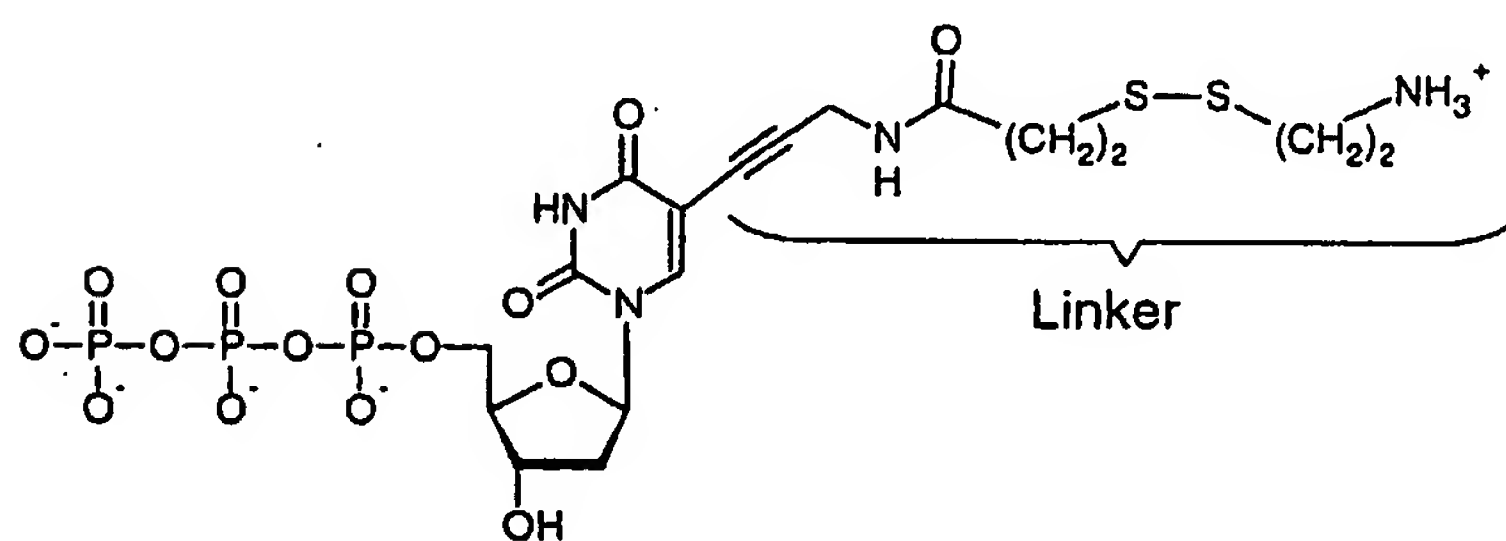


$X = \text{NH}, \text{O}, \text{S}$
 $X = \text{NH}, \text{O}, \text{S}$
 $Z = \text{NH}_2, \text{OH}$

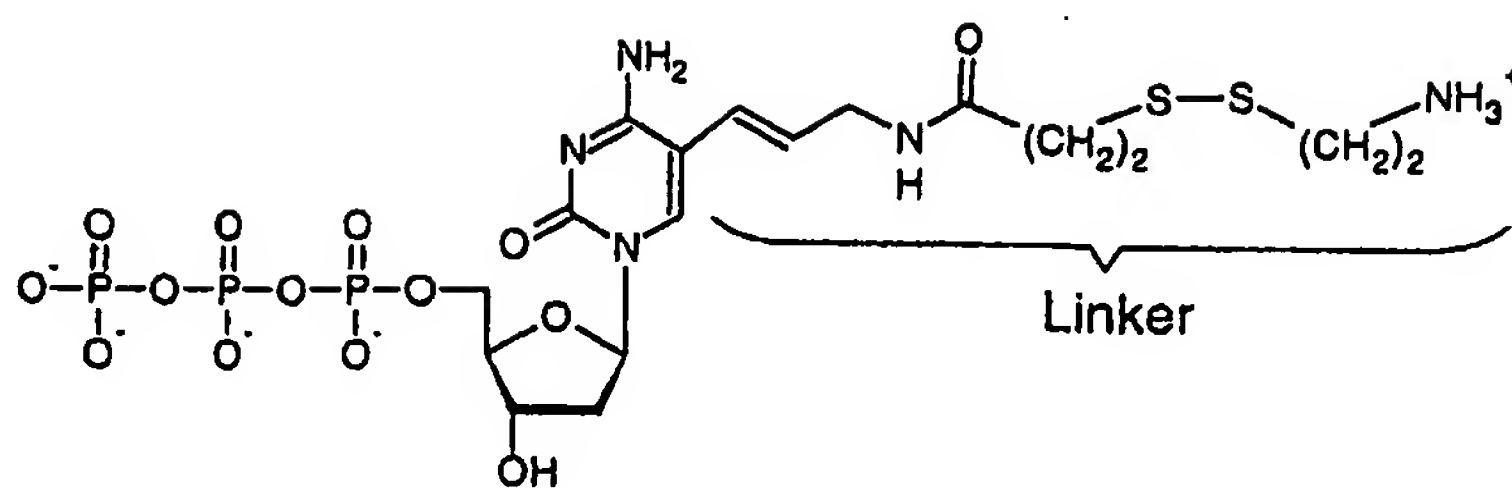
6f-1)



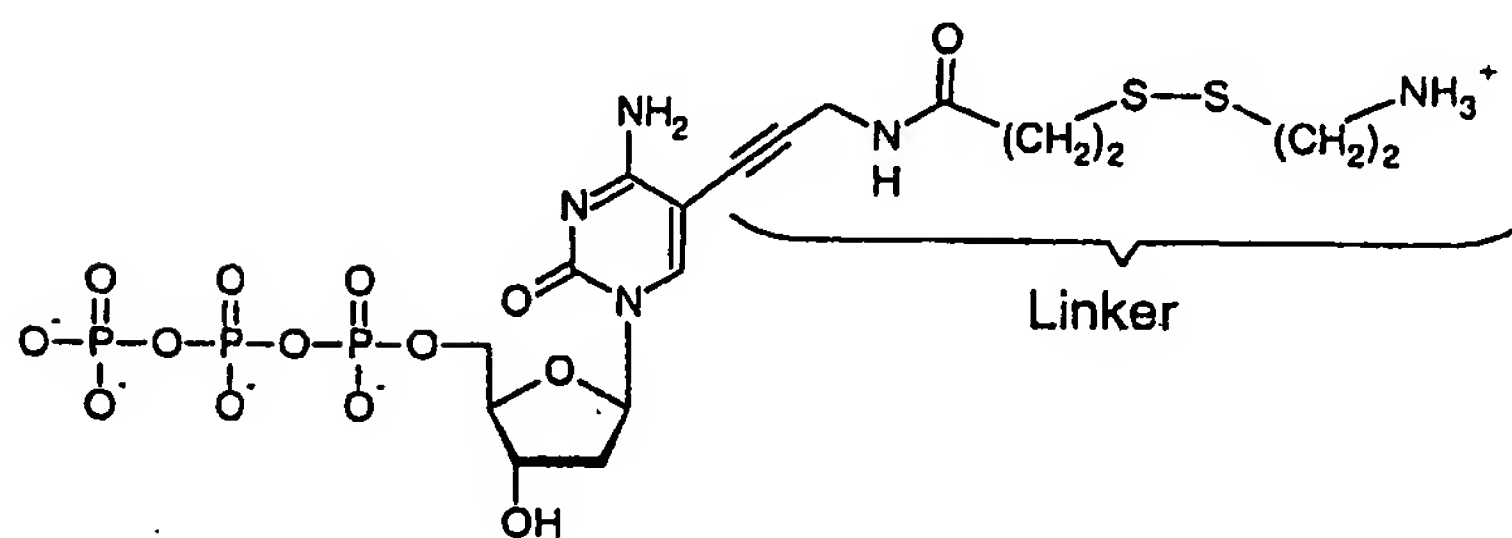
6f-2)



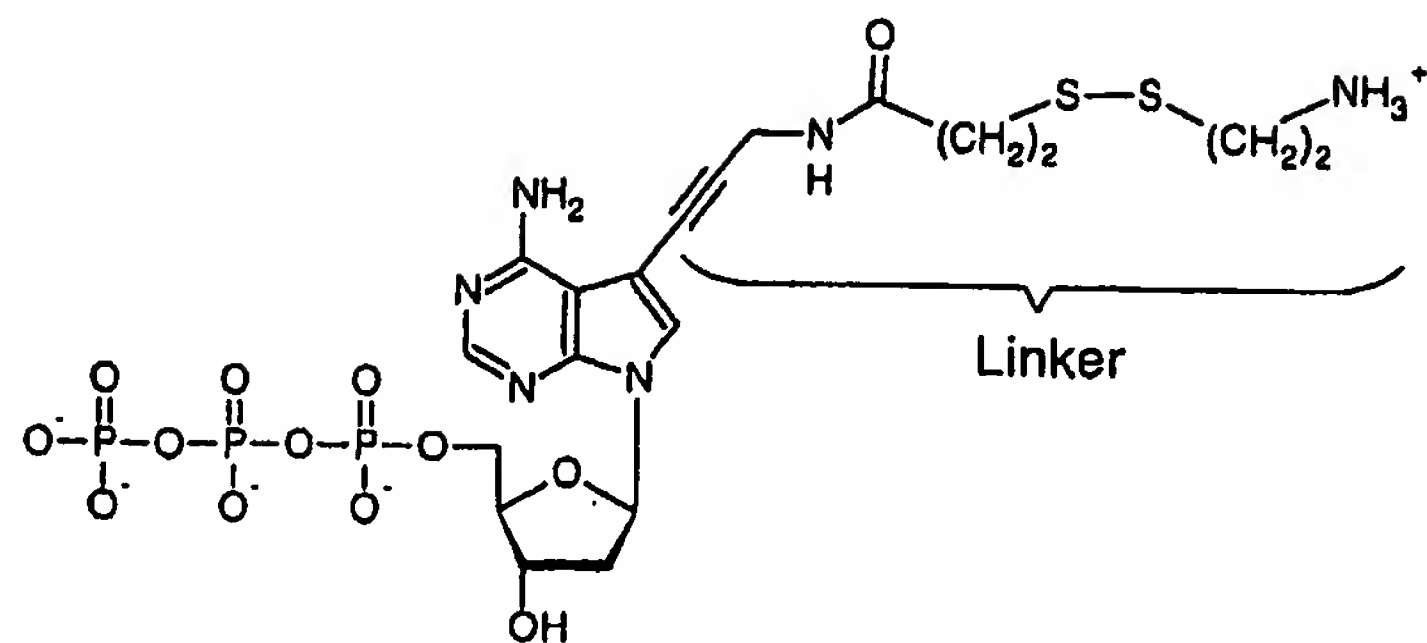
6f-3)



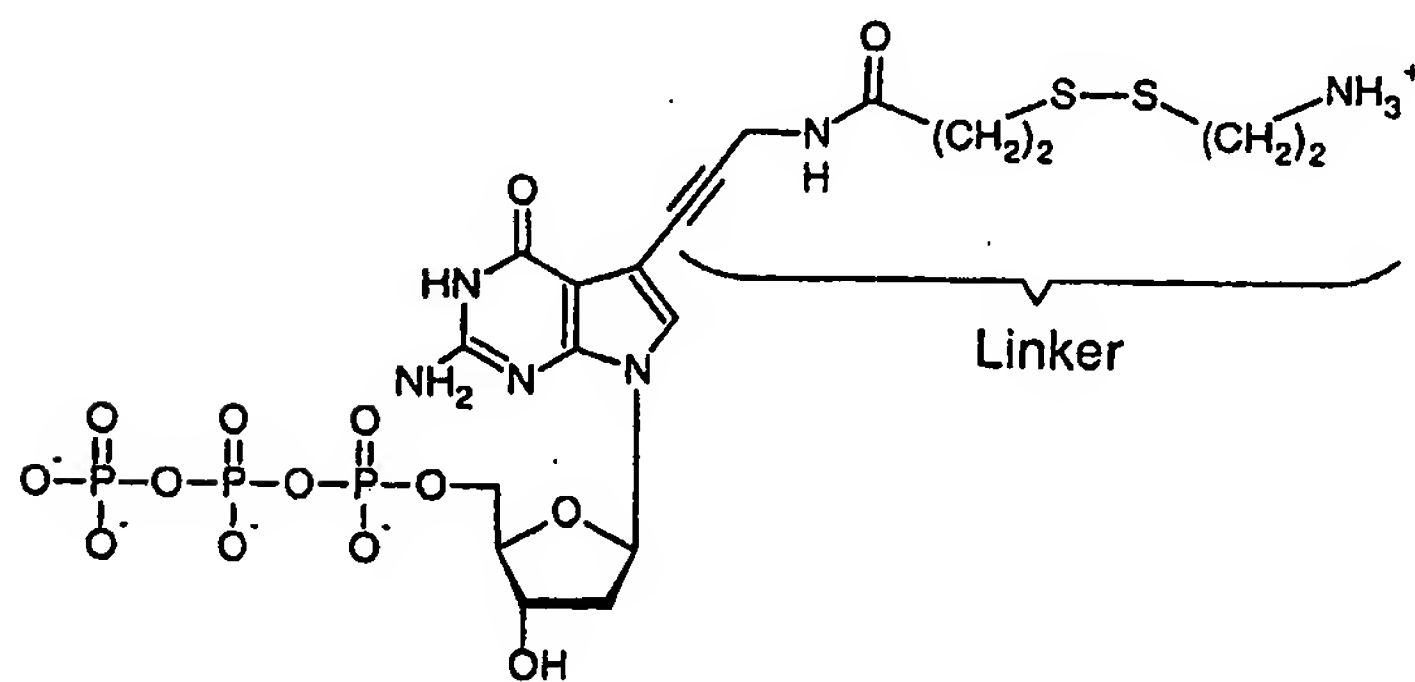
6f-4)



6g-1)



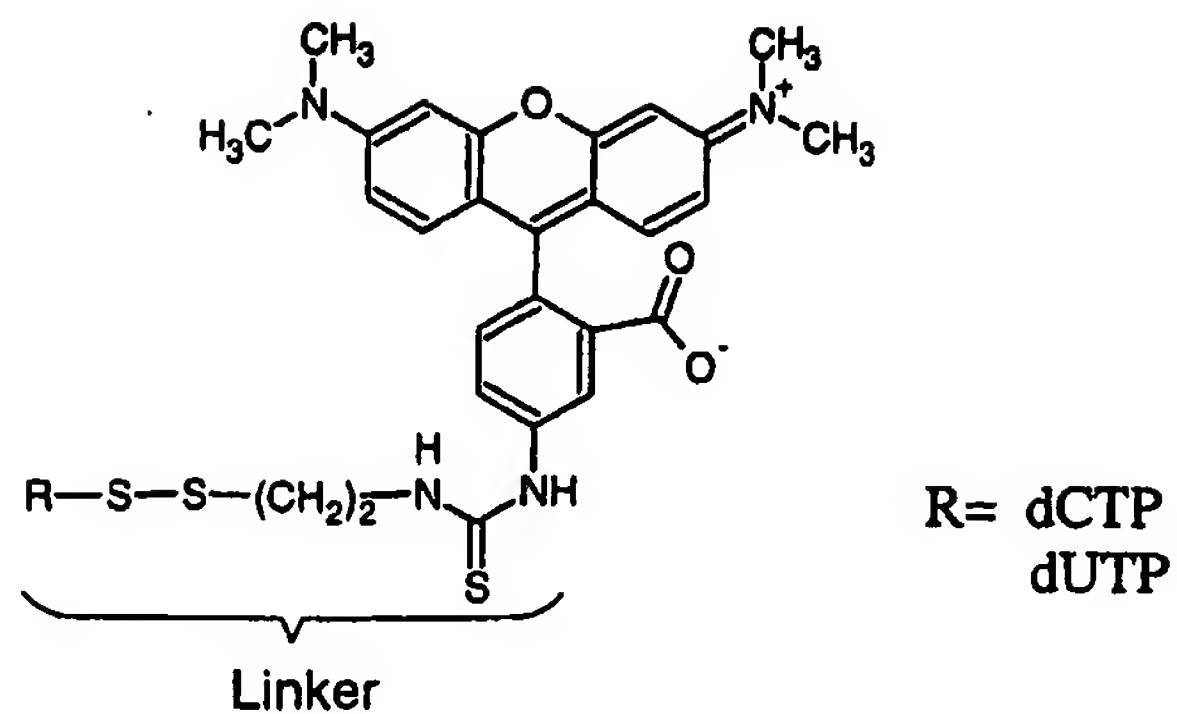
6g-2)



6h)

14/20

Fig. 6



6i)

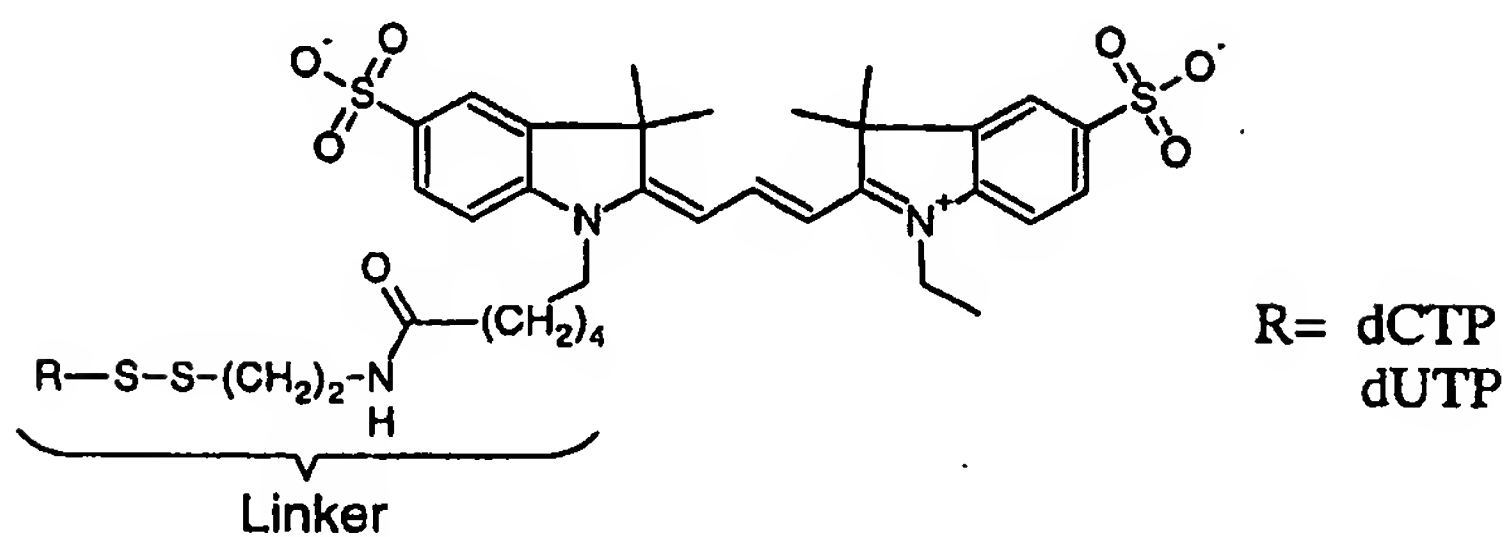
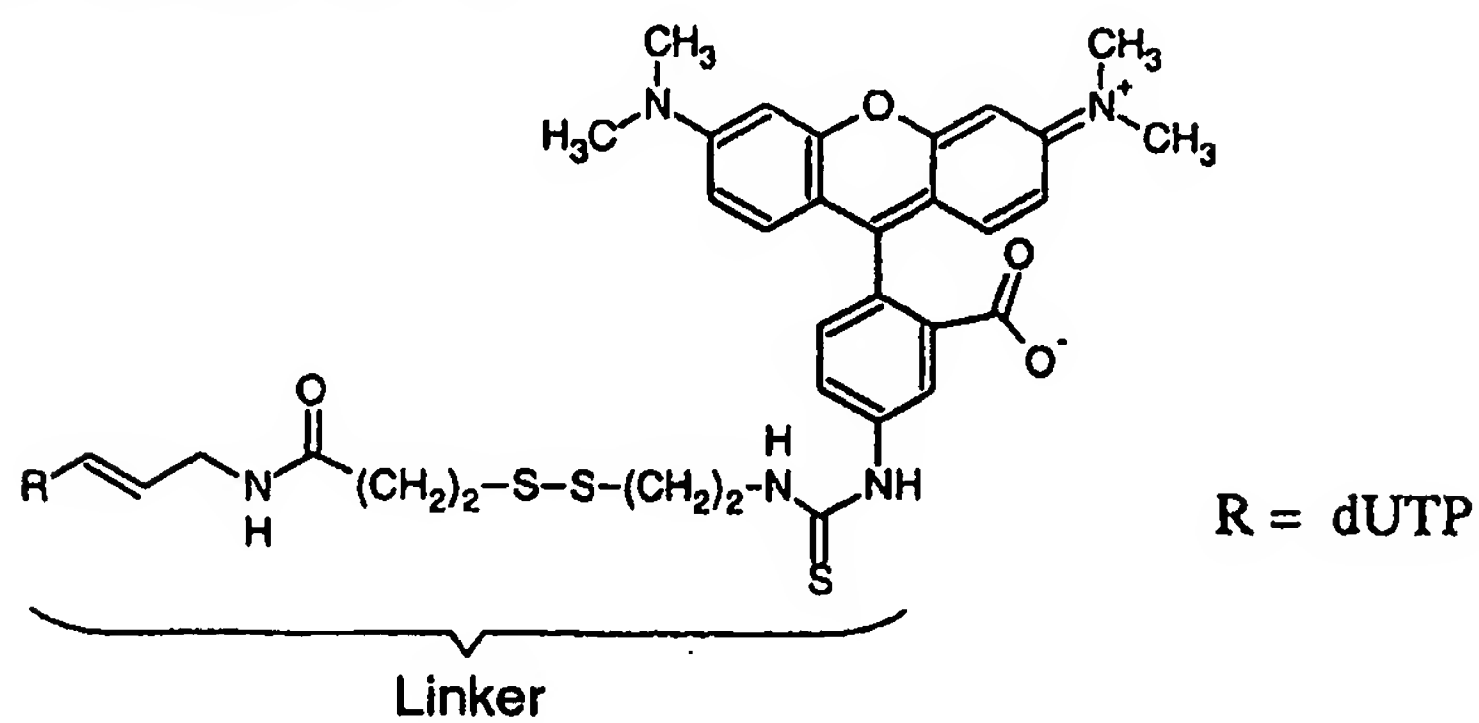
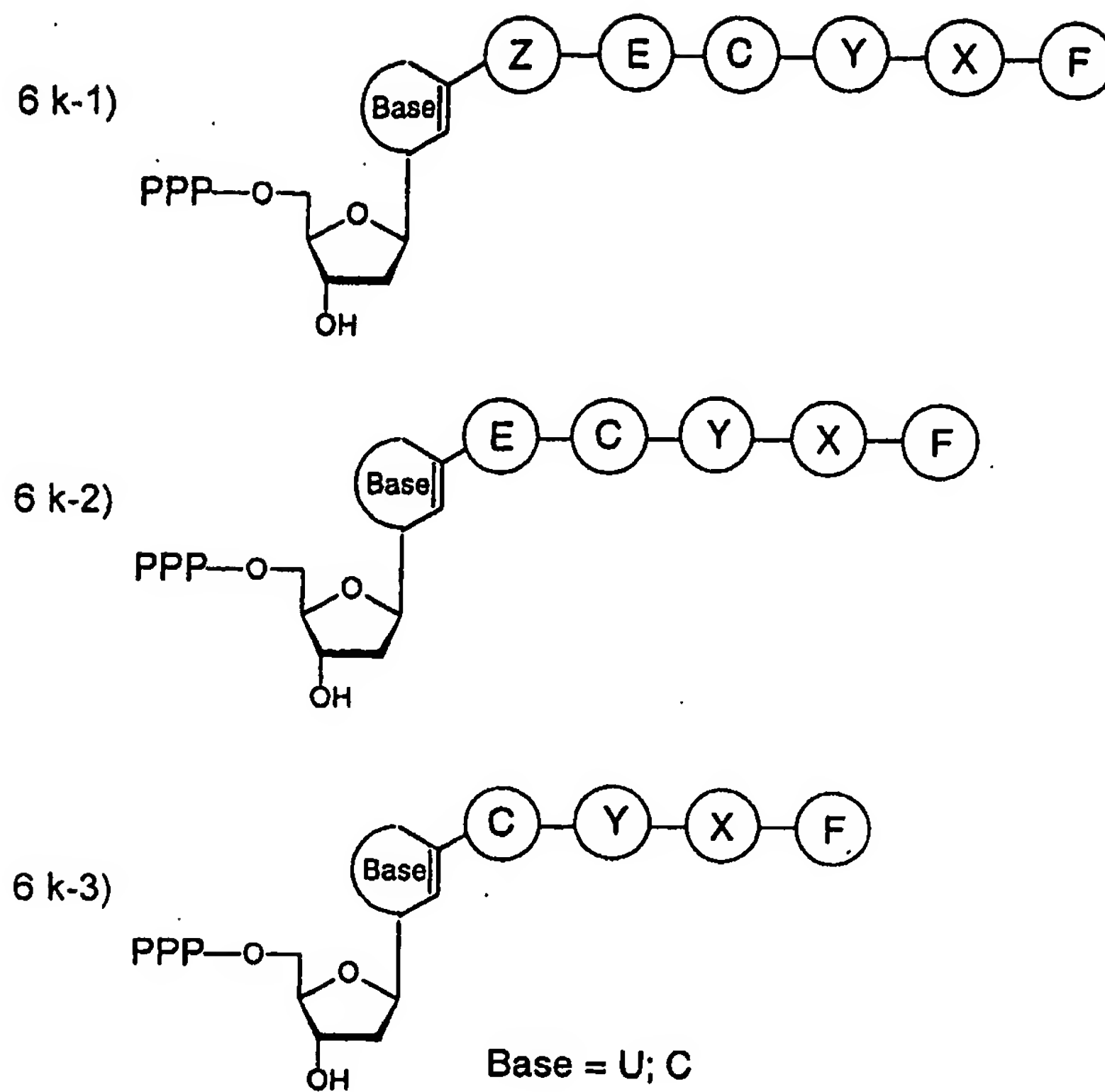
6j) dNTP-SS-TRITC (L 14)

Fig. 6



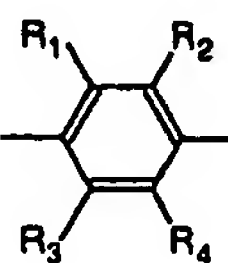
$R_1; R_2; R_3; R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-$

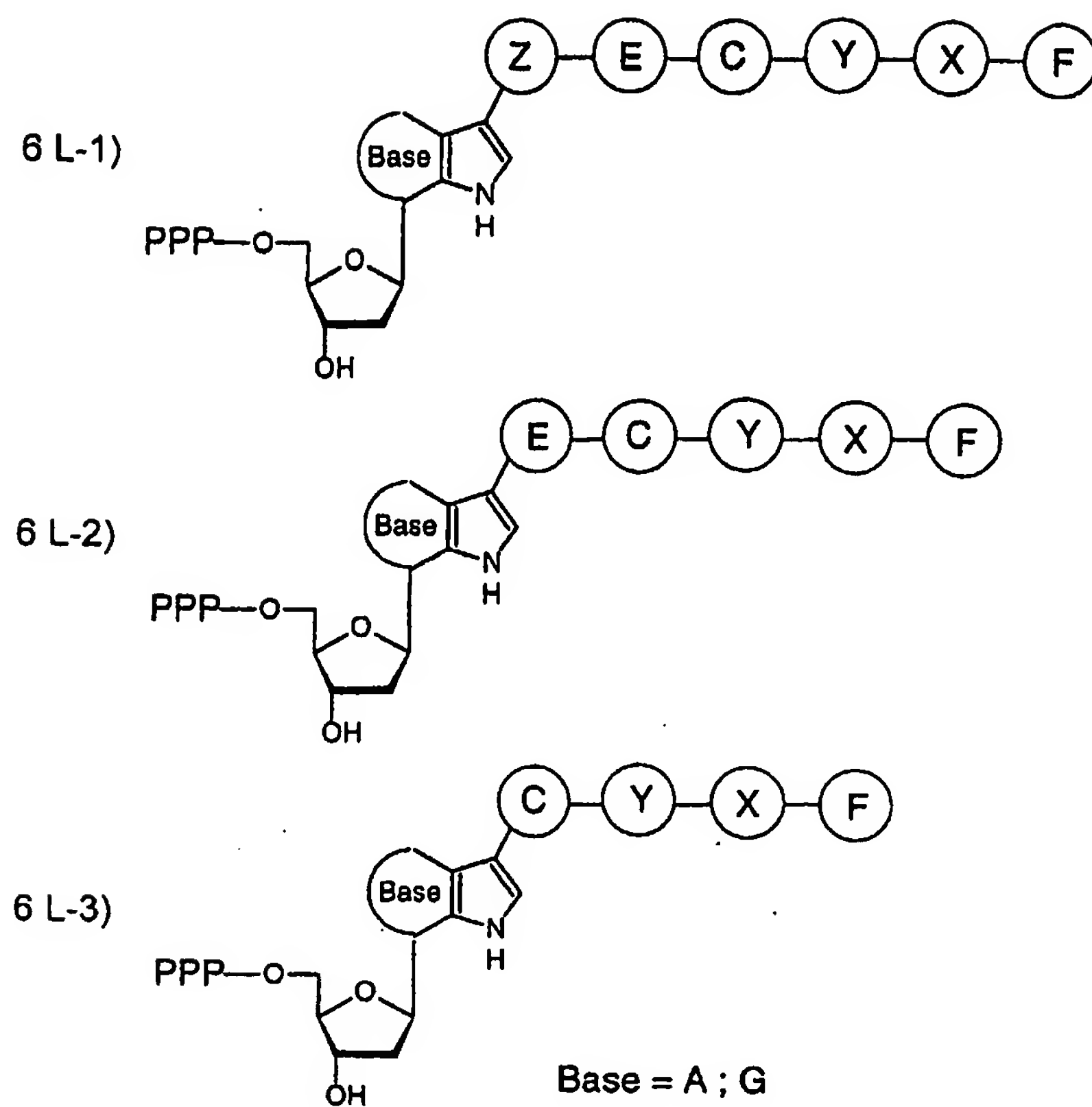
$Y = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-;$



$n_1; n_2; n_3; n_4 = 1 - 5$

F = Fluoreszenzfarbstoff

Fig. 6



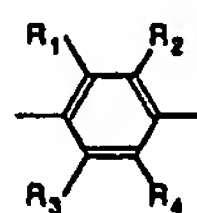
$R_1; R_2; R_3; R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-$

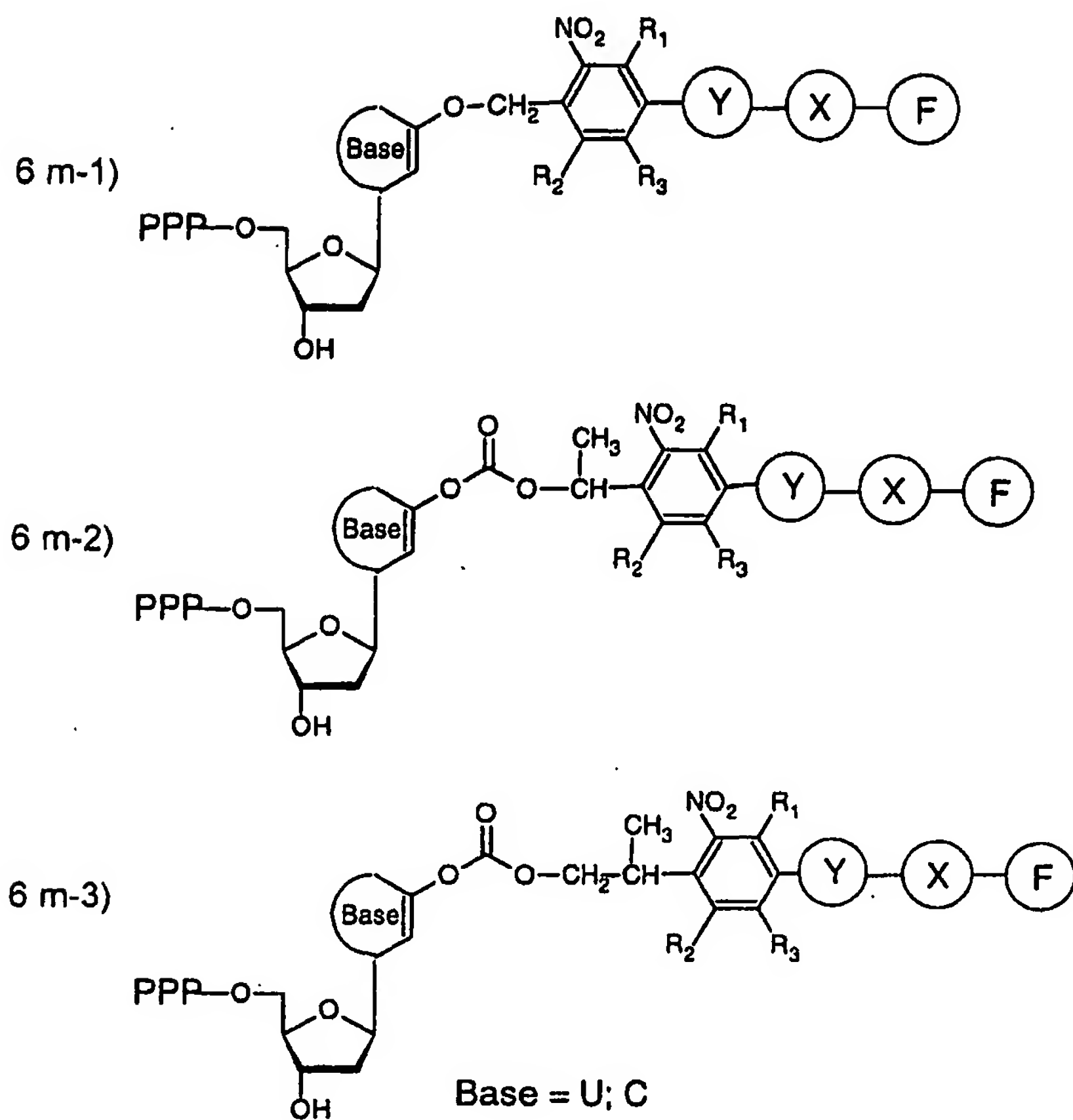
$Y = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-;$



$n_1; n_2; n_3; n_4 = 1 - 5$

F = Fluoreszenzfarbstoff

Fig. 6



$R_1; R_2; R_3; R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

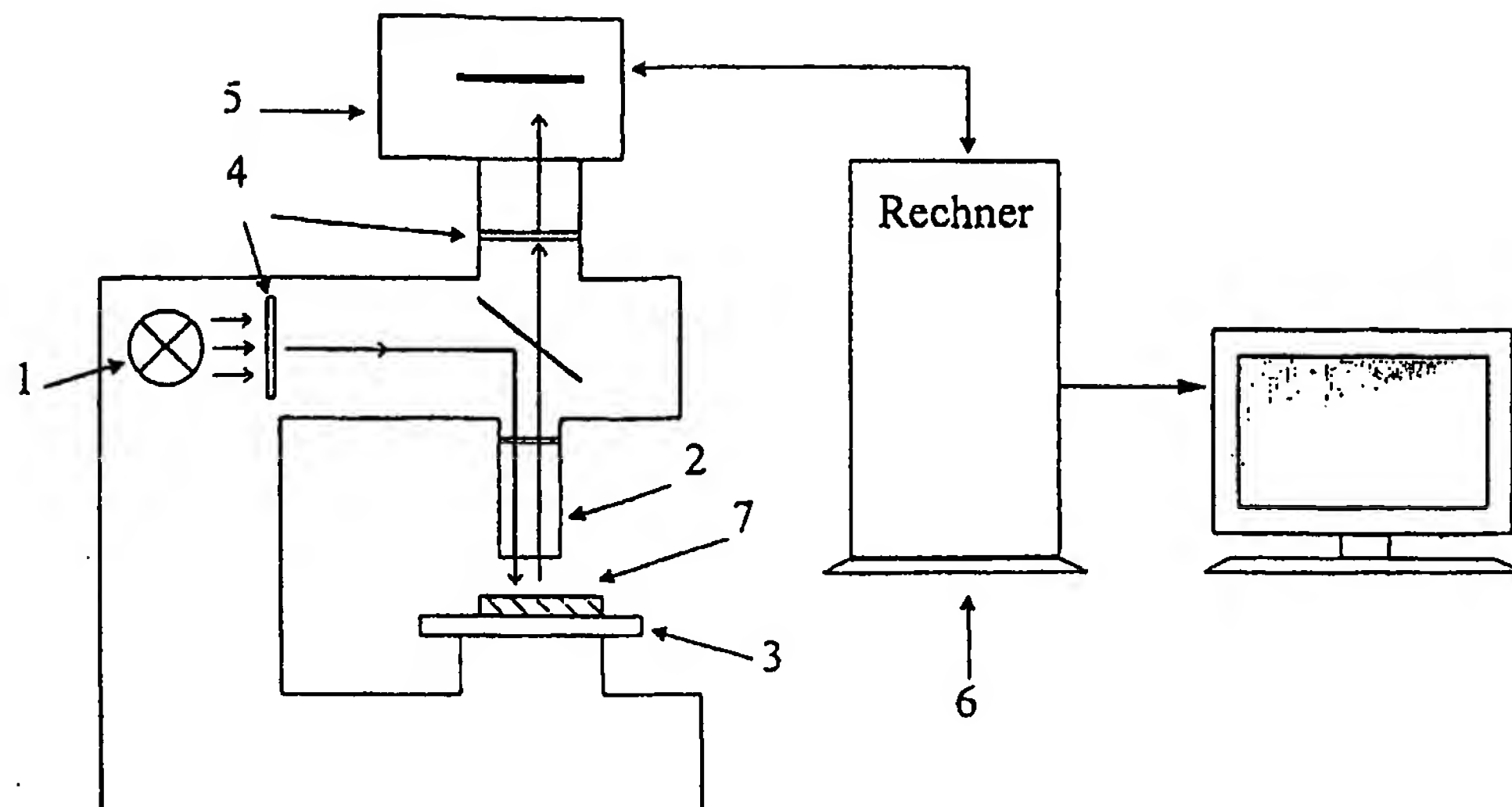
$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E; Y = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-$

$n_1; n_2; n_3; n_4 = 1 - 5$

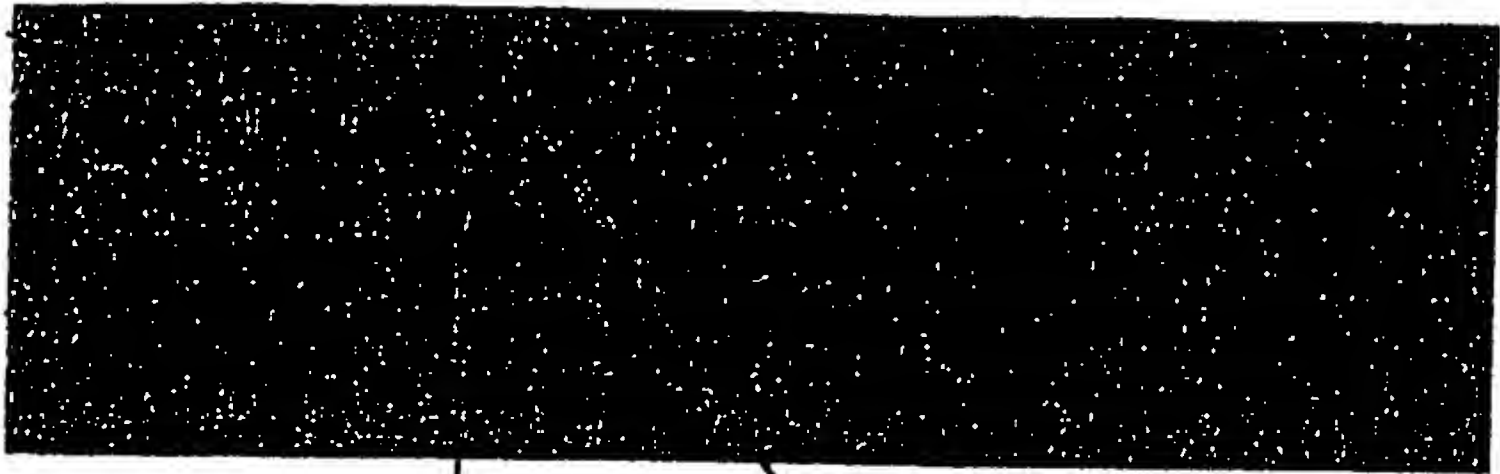
F = Fluoreszenzfarbstoff

**Legende:**

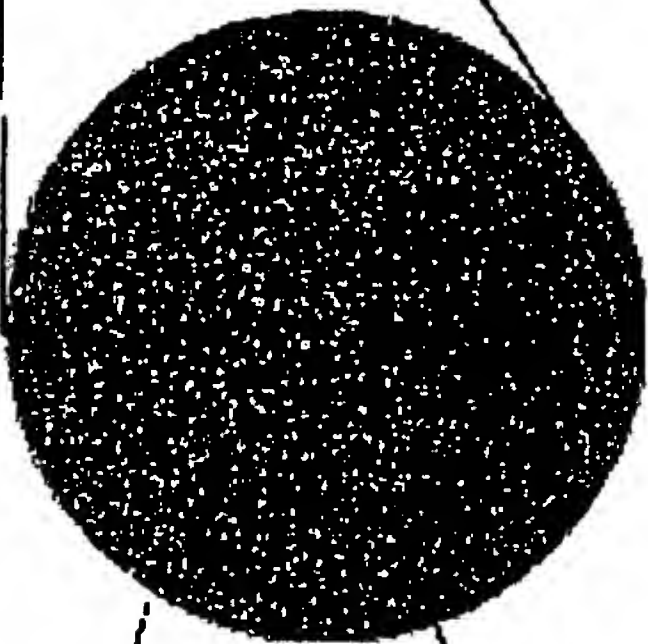
- Lichtquelle zur Anregung der Fluoreszenz (1)
- Lichtleitender Teil (2)
- Scanntisch (3)
- Vorrichtung zur Selektion von Spektren (4)
- Detektionsvorrichtung (5)
- Komputer mit Steuerungs- und Analysefunktionen (6)
- Oberfläche mit immobilisierten Genprodukte (7)

Fig. 8

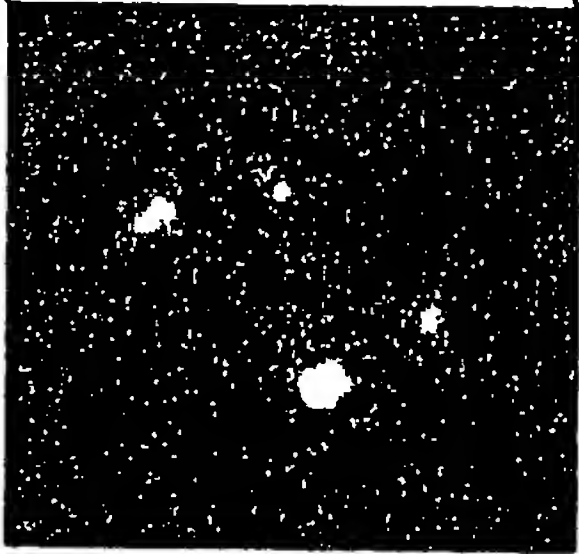
8a)



8b)

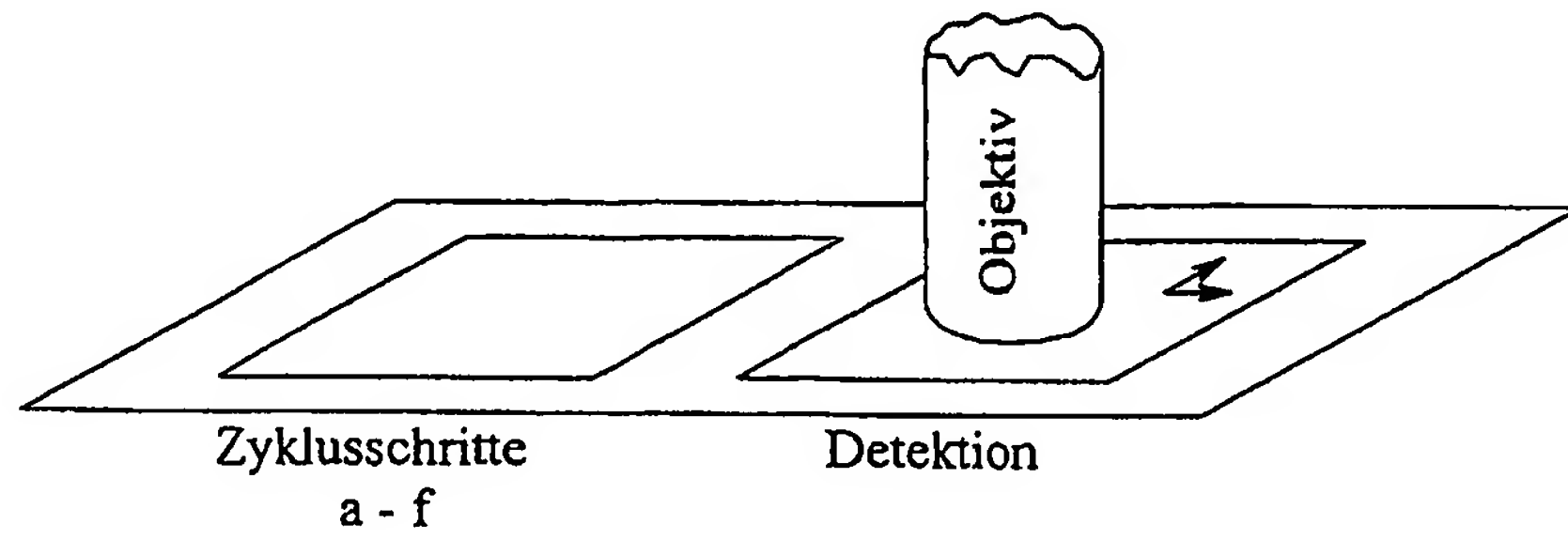


8c)



Sequenzierungsreaktion auf 2 getrennten Oberflächen.

Während eine Oberfläche gescannt wird, werden auf der andern Oberfläche alle anderen Zyklusschritte durchgeführt.



SEQUENZPROTOKOLL

<110> Dmitri Tcherkassov

<110> Prof.Dr. Walther Traut

<110> Dr. Englbert Bäuml

<110> Christian Hennig

<110> Norman Kock

<120> Verfahren zur Analyse der Genexpression

<130> P056483

<140>

<141>

<160> 2

<170> PatentIn Ver. 2.1

<210> 1

<211> 12

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: ermittelte Sequenz (Beispiel 4)

<400> 1

ACCAAAACAC CC

19

<210> 2

<211> 32

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Referenzsequenz (Beispiel 4)

<400> 2

ATCATCGTTC GAAATATCGA TCGCCTGATG CC

32

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.